DOCUMENT RESUME

ED 312 274 TM 013 999

AUTHOR	Owen, K.
TITLE	Test and Item Bias: The Suitability of the Junior Aptitude Tests as a Common Test Battery for White, Indian and Black Pupils in Standard 7. Report P-96.
INSTITUTION	Human Sciences Research Council, Pretoria (South Africa).
REPORT NO	ISBN-0-7969-0721-8
PUB DATE	89
NOTE	140p.
AVAILABLE FROM	Human Sciences Research Council, Private Bag X41, Pretoria 0001, Republic of South Africa.
PUB TYPE	Statistical Data (110) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Aptitude Tests; *Black Students; Comparative Analysis; Construct Validity; Cultural Differences; Foreign Countries; *Indians; Minority Groups; Predictive Validity; Psychometrics; Racial Differences; Secondary Education; *Secondary School Students; *Test Bias; Test Items; Test Validity; *White Students
IDENTIFIERS	*Junior Aptitude Tests (South Africa); Raven Progressive Matrices; South Africa

ABSTRACT

This study was undertaken to shed light on problems concerning the construction and use of a common test battery for various South African population groups. Specifically, the psychometric suitability for use with Indian and Black pupils, of the Junior Aptitude Tests (JAT), which were standardized for White pupils in Standard 7 was studied. The predictive validity and construct validity of the JAT and the test items were also investigated. The sample consisted of: (1) 1,056 White students from 30 schools; (2) 1,063 Indian students from 30 schools; and (3) 1,093 Black students from 28 schools. The students in the sample ranged in age from 13 to 20 years. Bias in the JAT was studied through comparison with results on the Raven Standard Progressive Matrices of J. C. Raven et al. (1977) and results on a reasoning test and biographical questionnaire developed for the study. Most tests of the JAT battery were found to be suitable for Indian students, but not for Black students. Differences between the JAT means for White and Black students were generally two standard deviations and more, and test reliabilities were low. Although construct bias was not apparent, bias was found in many test items. This study illustrates the difficulties involved in assessment of the diverse population groups in South Africa. A 109-item list of references is attached. Seven tables in Appendix A, 7 in Appendix B, and 60 graphs in Appendix C provide data to support the report. (SLD)

ED312274

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. J. PRINSLOO

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

INSTITUTE FOR PSYCHOLOGICAL AND EDUMETRIC RESEARCH P-96

TM013998

RGN·HSRC

Test and item bias: The suitability of the Junior Aptitude Tests as a common test battery for white, Indian and black pupils in Standard 7

K. Owen

BEST COPY AVAILABLE

Test and item bias: The suitability of the Junior Aptitude Tests as a common test battery for white, Indian and black pupils in Standard 7

Report P-96

Test and item bias: The suitability of the Junior Aptitude Tests as a common test battery for white, Indian and black pupils in Standard 7

K. Owen

Pretoria
Human Sciences Research Council
1989

K. Owen, D.Litt. et Phil.

Division for Aptitude Research
Institute for Psychological and Edumetric Research
Executive Director: Dr R.J. Prinsloo

ISBN 0 7969 0721 8

© Human Sciences Research Council, 1989

Published by the HSRC
Printed by Express Copy

CONTENTS

EKSERP IX

ABSTRACT X

1.0 INTRODUCTION 1

2.0 AIM 4

3.0 RESEARCH DESIGN 4

3.1 Samples 4
3.2 Measuring instruments 6
3.2.1 The Junior Aptitude Tests (JAT) 6
3.2.2 The Raven Standard Progressive Matrices (SPM) 10
3.3 Procedures, methods and techniques 11

4.0 GENERAL FUNCTIONING OF THE JAT FOR INDIAN AND BLACK PUPILS . 12

4.1 Means and standard deviations 12
4.2 Normality of the score distributions 14
4.3 Testing time 16

5.0 BIAS IN THE PREDICTIVE VALIDITY OF THE JAT 23

5.1 Differential and single-group validity 23
5.2 The slope, intercept and standard error of estimate of the regression line 26
5.3 The use of a common regression line 28

6.0 BIAS IN THE CONSTRUCT VALIDITY OF THE JAT 29

6.1 Definitions of bias in the construct validity of a test . 29
6.2 Construct validity of the JAT for white, Indian and black pupils 30
6.2.1 Test reliabilities 30
6.2.2 Rank order of item difficulties 33
6.2.3 Factorial similarity 33

7.0 ITEM BIAS IN THE JAT 40

7.1 Definitions of item bias 40
7.2 The detection of item bias 41
7.2.1 The chi-square 42
7.2.2 The item response theory 45
7.3 Item bias in the tests of the JAT 48
7.3.1 JAT 1: Classification 48
7.3.2 JAT 2: Reasoning 51
7.3.3 JAT 3: Number Ability 56
7.3.4 JAT 4: Synonyms 57
7.3.5 JAT 5: Comparison 57
7.3.6 JAT 6: (2-D) and JAT 7 :(3-D) 57
7.3.7 JAT 8: Memory (Paragraph) and JAT 9: Memory (Words and Symbols) 59

7.3.8 JAT 10: Mechanical Insight 59
7.4 Summary 59

8.0 DISCUSSION AND RECOMMENDATIONS 61

8.1 The suitability of the JAT for Indian and black pupils . 61
8.2 Bias: predictive validity, construct validity and test items 62
8.3 Recommendations 68

REFERÉNCES 70

APPENDIX A Percentage of testees who selected a particular option . 77

APPENDIX B Bias indices 85

APPENDIX C Item characteristic curves (ICCs) for JAT1 and JAT2 . . 93

LIST OF TABLES

TABLE 3.1 WHITE, INDIAN AND BLACK SAMPLES: NUMBER OF SCHOOLS AND PUPILS (STANDARD 7) 5

TABLE 3.2 MEAN ($\bar{X}$) AND STANDARD DEVIATION(S) OF THE AGE (IN YEARS) OF THE THREE POPULATION GROUPS AND DISTRIBUTION OF PUPILS BY AGE GROUP 6

TABLE 3.3 TESTING TIMES FOR THE JAT 10

TABLE 4.1 THE MEANS ($\bar{X}$) AND STANDARD DEVIATIONS (S) OF THE JAT FOR WHITE, INDIAN AND BLACK PUPILS AND THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE GROUPS (HOTELLING T^2 AND t TESTS) 13

TABLE 4.2 DIFFERENCES BETWEEN THE TEST MEANS OF WHITE AND INDIAN AND WHITE AND BLACK PUPILS IN TERMS OF THE STANDARD DEVIATION OF THE WHITES 14

TABLE 4.3 SKEWNESS (SK), KURTOSIS (KU) AND RELIABILITY (K-R21) OF THE TEST SCORES OF THE JAT (STANDARD TESTING TIME) FOR WHITE, INDIAN AND BLACK PUPILS 15

TABLE 4.4 PROPORTION OF THE TESTEES WHO HAD REACHED A TEST ITEM IN A CERTAIN TESTING TIME (T) 17

TABLE 4.5 TEST MEANS OF THE WHITE, INDIAN AND BLACK PUPILS WITH REGARD TO A CERTAIN TESTING TIME(T) 21

TABLE 4.6 MEAN TEST SCORES ON TESTS 1, 2, 3, 5 AND 6 OF THE JAT FOR TESTEES WHO COMPLETED A TEST IN TIME 1 (GROUP A), TIME 2 (GROUP B) AND TIME 3 (GROUP C) 21

TABLE 4.7 SUMMARY OF ANALYSIS OF VARIANCE OF THE SCORES ON TESTS 1, 2 AND 3 OF THE JAT WITH REGARD TO POPULATION AND TIME GROUP 22

TABLE 5.1 PREDICTIVE VALIDITY: CORRELATION COEFFICIENTS (r_{xy}), SIGNIFICANCE OF THE DIFFERENCES BETWEEN POPULATION GROUPS (Z), CONFIDENCE INTERVALS AND CORRECTION FOR ATTENUATION ($k r_{xy}$) FOR THE JAT 1 AND THE SPM 25

TABLE 5.2 JAT 1 AND SPM: SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE POPULATION GROUPS IN THE SLOPES, INTERCEPTS AND STANDARD ERRORS OF ESTIMATE OF THE REGRESSION LINES 27

TABLE 5.3 DEVIATION (D) IN STANDARD DEVIATION UNITS OF PREDICTED MEAN ($\bar{\bar{Y}}$) FROM OBTAINED MEAN ($\bar{Y}$) ON THE SPM FOR INDIAN AND BLACK PUPILS 28

TABLE 6.1 SIGNIFICANCE OF THE DIFFERENCES (F) BETWEEN THE TEST RELIABILITIES (K-R20) OF THE POPULATION GROUPS WITH REGARD TO THE JAT AND SPM 32

TABLE 6.2 RANK ORDER CORRELATION (SPEARMAN RHO) BETWEEN THE ITEM DIFFICULTY VALUES FOR THE JAT FOR WHITE/INDIAN AND WHITE/BLACK PUPILS, AND PROPORTION OF GROUPS X ITEMS INTERACTION VARIANCE ($1-RHO^2$) 33

TABLE 6.3 INTERCORRELATIONS OF THE TEN TEST SCORES OF THE JAT FOR WHITE, INDIAN AND BLACK PUPILS 34

TABLE 6.4 PRINCIPAL COMPONENT ANALYSIS OF THE TESTS OF THE JAT FOR WHITE, INDIAN AND BLACK PUPILS 35

TABLE 6.5 FACTOR LOADINGS AND COMMUNALITIES (h^2) OF THE TESTS OF THE JAT FOR WHITE, INDIAN AND BLACK PUPILS 37

TABLE 6.6 COEFFICIENTS OF CONGRUENCE FOR A THREE-FACTOR SOLUTION FOR WHITE, INDIAN AND BLACK PUPILS 38

TABLE 6.7 RESIDUAL MATRICES FOR WHITE, INDIAN AND BLACK PUPILS 39

TABLE 7.1 ABILITY GROUPS ACCORDING TO RANGE OF TOTAL SCORES ON THE SPM, AND NUMBER OF TESTEES IN EACH ABILITY GROUP . . 44

TABLE 7.2 JAT 2: REASONING: RANGE OF CHI-SQUARE (X^2_3) AND ITEM DIFFICULTY VALUES (p) ACCORDING TO ITEM TYPE 54

TABLE 7.3 SUMMARY OF THE PROBABLE CAUSES OF ABERRANT TEST ITEMS 60

LIST OF FIGURES

FIGURE 4.1 REPRESENTATION OF THE PROPORTION OF THE WHITE (W), INDIAN (I) AND BLACK (B) PUPILS WHO HAD REACHED THE LAST ITEM IN A TEST IN A CERTAIN TESTING TIME 19

FIGURE 7.1 JAT 1: SCATTERGRAM OF ITEM DIFFICULTY VALUES FOR WHITE/INDIAN AND WHITE/BLACK 50

FIGURE 7.2 JAT 2: SCATTERGRAM OF ITEM DIFFICULTY VALUES FOR WHITE/INDIAN AND WHITE/BLACK 52

EKSERP

Hierdie ondersoek is onderneem ten einde meer lig te werp op die vraagstukke betrokke by die opstelling en gebruik van 'n gesamentlike toetsbattery vir die verskillende bevolkingsgroepe in Suid-Afrika.

Die oogmerke met die studie is eerstens om die psigometriese geskiktheid van die Junior Aanlegtoetse(JAT), wat vir blanke skoliere gestandaardiseer is, ook vir Indiër-en swart skoliere in standerd 7 te ondersoek en tweedens om sydigheid - en die moontlike oorsake daarvan - ten opsigte van die voorspellingsgeldigheid en konstrukgeldigheid van die toetsbattery en in toetsitems te identifiseer.

Daar is bevind dat meeste toetse van die JAT geskik is vir die Indiërleerlinge en hoewel hul toetsgemiddeldes sowat een standaardafwyking laer as dié van die blankes is, is die toetsbetroubaarhede soortgelyk vir die twee groepe. Die JAT funksioneer egter onbevredigend vir die swart leerlinge: verskille in toetsgemiddeldes tussen die blanke en swart leerlinge is meestal twee standaardafwykings en groter en die toetsbetroubaarhede is laag.

Sydigheid ten opsigte van voorspellingsgeldigheid is vir JAT 1 (Klassifikasie), met die Raven SPM as kriterium, ondersoek. Dit blyk dat hoewel differensiële en enkelgroepgeldigheid grootliks afwesig is, sydigheid voorkom in die helling, afsnit en standaardfout van die regressielyn. Die kriteriumprestasie van die swart leerlinge word oorvoorspel deur die regressievergelyking van die blanke leerlinge. Faktorontleding het getoon dat die JAT waarskynlik dieselfde sielkundige konstrukte by die drie bevolkingsgroepe meet. Konstruksydigheid kom dus nie voor nie.

Die sydigheidsindekse(Scheunemantegniek, volle chi-kwadraat en die drieparameter logistiese model van die itemresponsteorie) toon talle sydige items. 'n Onderskeid word gemaak tussen ware sydigheid wat sy oorsprong in die item het soos spesifieke taalkennis of bekendheid met voorwerpe, en skynsydigheid wat uit die toetsgedrag van die toetsling spruit, byvoorbeeld 'n neiging tot assosiatiewe konseptualisering en 'n gebrek aan 'n logiese ingesteldheid.

Implikasies van die ondersoek vir verdere navorsing word aangedui.

ABSTRACT

This study was undertaken to shed more light on the problems concerning the construction and use of a common test battery for the various population groups of South Africa.

The objectives of this study are firstly to investigate the suitability of the Junior Aptitude Tests (JAT), which were standardized for white pupils, for Indian and black pupils in Standard 7, and secondly to identify bias - and its possible causes - concerning the predictive validity and construct validity of the test battery and test items.

It was found that most JAT tests were suitable for Indian pupils and although their means were one standard deviation lower than those for whites, the test reliabilities for the two groups were similar. The JAT was not satisfactory in the case of black pupils. Differences between the test means for white and black pupils were mostly two standard deviations and more, and the test reliabilities were low.

Bias concerning the predictive validity of the JAT 1 (Classification) was investigated by making use of the Raven SPM as criterion. It appeared that although differential and single-group validity was mainly absent, bias was detected in the slope, intercept and standard error of the regression line. The criterion achievement of the black pupils was overpredicted by the regression equation of white pupils. Factor analysis indicated that the JAT probably measured the same psychological constructs in the three population groups. Construct bias therefore did not appear.

The bias indices(Scheuneman technique, full chi-square and the three parameter logistical model of item response theory) indicated many biased items. A distinction was made between true bias which originated in the test item itself, for example specific language knowledge or familiarity with objects, and apparent bias resulting from the pupil's test behaviour, for example a tendency to associative conceptualization and an absence of a logical attitude.

The implications of this investigation for further research are indicated.

1.0 INTRODUCTION

The suitability of a psychometric test for a group that did not form part of the standardization group is closely related to the problem of bias. According to Reynolds and Brown (1984) this matter came to light mainly because of the nature of psychological processes and the measurement of such processes. Psychological processes cannot be observed or measured directly and the way a process works therefore has to be derived indirectly from behaviour. Psychologists have reached consensus on very few of these deductions or hypothetical constructs and it is therefore understandable that intelligence or intellectual ability, which represents what is probably one of the most complex processes in psychology, can stimulate the interest of both experts and laymen. The criticism - including that of bias - levelled against psychometric tests such as intelligence and aptitude tests should also be seen against this broad background. Complaints about test bias are particularly heard from minority groups in the USA and elsewhere and according to them certain tests are more suitable for the group that formed the largest component of the standardization sample. Whether such tests are indeed biased and harmful to the interests of minority groups is one of the questions that have to be answered empirically.

To limit the potential influence of cultural factors on test performance Cattell in 1940 suggested a "culture free" intelligence test. Owing to the misunderstanding and misrepresentation caused by the term "culture free" the name was later changed to "culture fair" intelligence tests. At that stage it was not clear whether there was really a link between culture and test performance. The first systematic investigation in the USA with regard to cultural bias in psychometric tests was conducted in 1945 by Allison Davies, a sociologist, and Kenneth Eells, a psychologist (Jensen 1980). These researchers regarded cultural bias in test items as only one of several possible causes, including hereditary traits, developmental factors, motivation factors, work habits, skill in writing tests, etc., which contribute towards the average IQ differences between various culture groups.

According to Jensen (1980) the confusion in the literature regarding terms such as "bias" and "unfairness" that developed after the pioneer work by Eells can be attributed to a lack of conceptual clarity regarding the meaning of these terms. For example, "cultural bias" became a popular cliché in the explanation of race and social differences found in respect of performances in intelligence tests. Although there was much conjecture on the subject, researchers did not try to determine the real nature of the difference in test scores or to establish objective criteria for determining bias in a specific test.

During the sixties the situation in the USA became more pressing and concern developed over the fact that the test scores of blacks and other minority groups (the "culturally deprived") were generally lower than those of white Americans. These differences were found in a wide variety of tests, for example intelligence, scholastic aptitude and performance tests. Because the tests revealed differences they were condemned as being culturally biased against blacks by sociologists, anthropologists, educationists and other critics who actually were outside the field of psychometrics. As a result pressure groups formed that advocated the total abolition of psychological tests. However responsible persons and institutions realized that instead of abolishing

tests attention should rather be paid to the fair use of tests on the one hand and the possible bias of such tests on the other.

During the seventies pyschometricians started to study concepts such as test bias and test fairness in a more systematic way. Not only did this result in greater clarity in the definitions and the terminology, but it also led to a great deal of research.

The term *bias* has been used quite a few times so far and a brief definition of the concept is probably in order (it will be discussed in more detail later). In mathematical statistics the term bias refers to a systematic underestimate or overestimate of a population parameter by a statistic that is based on a sample from that population. In psychometrics the term refers to systematic errors in the *predictive validity* or the *construct validity* of the test scores of individuals that are associated with the group membership of those individuals (Jensen 1980).

The above two aspects of test scores, namely predictive validity and construct validity, represent the two main fields of research on test bias. At first glance the two fields are completely divergent and have little in common with regard to methods of investigation. However the two approaches or fields do *not* represent different concepts of bias and, as Raju and Normand (1985) indicated, in certain cases the same method of analysis can be used for both approaches.

Research conducted in the USA during the past 10 to 15 years on congnitive tests for minority groups and the possible bias of such tests can be divided into four main categories. In the first category of experiments, which also represents most of the research, the *predictive validity* of the tests, in other words the correlation between the test and criterion scores, was studied. This type of research is known as "test bias in the presence of an external criterion". An equivalent term that is often found in the literature on the subject is "selection bias".

The second category of experiments investigates the way tests are used and particularly *fair selection models*. Although this type of research is not directly related to test bias (a biased test can be used in a fair or unfair manner and the same also applies to an unbiased test) it is nevertheless of interest in the broad context of the testing and selection of minority groups.

The third category is concerned with *construct validity* of a test, in other words what a test is supposed to measure. Because a hypothetical construct that can seldom be verified in practice with the aid of an external criterion is usually involved, this type of research is known as "test bias in the absence of an external criterion". When studying this form of bias the internal characteristics of a test such as reliability, internal consistency and the difficulty value of its items are studied. On the basis of these characteristics it can be decided whether the construct validity of the test is adequate for a specific population group. It is also important to note that although *item bias* is often regarded as a separate field of study in the literature, among other things because specific methods had to be developed to study bias in this regard, item bias nevertheless forms part of the construct approach to test bias.

The fourth category includes all the experimental work done to develop methods for studying bias and to compare the different methods with one another with the aid of real or simulated data in order to find the "best" method for identifying bias.

The question of whether to use common or separate psychometric tests for the various population groups in South Africa is extremely relevant at this juncture. Not only do some groups resist the use of separate tests (according to them this step only emphasizes their "separateness" or "uniqueness"), but the constitutional changes that have been implemented since 1983 as well as other processes and changes that came into effect earlier have had a ripple effect that can be observed in many fields.

The two largest constructors of psychometric tests in South Africa, namely the Institute for Psychological and Edumetric Research (IPER) of the Human Sciences Research Council (HSRC) and the National Institute for Personnel Research (NIPR), which was incorporated into the HSRC in 1984, have so far constructed mainly separate but similar tests for the various population groups. Because every population group had its own education department with its own particular needs, separate tests had to be constructed. At postschool level too there was little specific need for common tests because the various groups did not compete with each other.

This image of "peaceful separateness" was buried officially on 12 September 1985 when State President P.W. Botha issued the following statement to the press (Directorate: Liaison and Information 1985):

> *As far as job reservation is concerned, all forms of statutory job reservation have been abolished in accordance with the Government's declared policy, while the principle of abolishing job reservation in the mining industry was accepted in 1981 already. In the Public Service, too, it is found that trained persons of all population groups are increasingly being appointed to positions previously held exclusively by Whites* (p. 7).

The changing industrial situation in South Africa and the abolition of job reservation, which meant that in many cases the members of the various population groups could now compete for the same jobs, posed the following problem for the recruiting and selection personnel at the various companies: On what basis could a decision be made as to which candidate was the best or most suitable if all the candidates did not do *the same* psychometric tests? If all of them did the same tests (for instance tests that had been developed for whites), would one group not perhaps benefit at the expense of the other groups? Just as in the USA and in other plural societies, the problem of bias in psychometric tests has arisen in the RSA as well.

In spite of the sharp differences of opinion in certain academic circles in South Africa with regard to the problem of common/separate tests (some academics are completely opposed to separate tests, while others feel that common tests are not feasible as they will not be able to comply with strict psychometric requirements), it is clear from the above that the construction and use of common psychometric tests in South Africa is unavoidable. These common tests will naturally cover a wide

range of abilities, characteristics, skills and fields of knowledge and it is therefore extremely important for research to be conducted on test bias.

2.0 AIM

This study has two aims. The first is to investigate the psychometric suitability for Indian and black pupils of a test that was standardized for white pupils in South Africa (it was not possible to involve coloured pupils as well). The second aim is to study the two main aspects of test bias in cognitive tests, namely bias in respect of predictive validity and construct validity. As far as item bias (as part of the construct approach) is concerned, the intention is to identify not only the characteristics and particular content that contribute towards bias, but also (and even more important) to search for factors on the basis of which the reasons for bias can be determined.

The intention is therefore to throw more light on the problems experienced when the same test is used for different population groups.

3.0 RESEARCH DESIGN

In this section attention is given to the samples (testees) involved in the investigation, the measuring instruments used, the procedures followed in collecting the data and the methods and techniques used to analyse the data.

3.1 SAMPLES

The samples were directly related to the basic problem of the investigation, namely the extent to which a test - more specifically the Junior Aptitude Tests (JAT) - that had been constructed for white pupils was suitable for other population groups in South Africa as well.

The assumption that the need for common tests for the different population groups was greater in urban than in non-urban areas, mainly as a result of better job opportunities, led to the selection of only urban samples. A further assumption in the selection of these samples was that urban pupils, irrespective of population group, were exposed to the influence of the same large business centres and other economic activities and communication media such as newspapers, magazines, films, radio and television and that they were therefore more equal in respect of the variables that could influence test performance.

Another consideration in the selection of the samples was that, as far as possible, the testees should be on the same level with regard to academic qualifications. It was therefore decided to involve only Standard 7 pupils. This step does not imply that all pupils in the same standard at school necessarily receive the same standard of instruction or have the same knowledge. However it can be assumed that all the pupils have at least a minimum amount of school experience in common. Although the tests of the JAT are not based on the school syllabus or demand a lot of school experience, the tasks presented in the tests cannot be divorced completely from experiences that pupils come across formally at school or informally in daily life.

The size of the sample was determined to a certain extent by the statistical techniques that were used to analyse the data. For instance techniques that are based on item response theory require a minimum of approximately 1 000 pupils per group in order to make a reasonably accurate estimate of the item parameters. It was therefore decided to test at least that number of pupils in each population group.

The samples for the white, Indian and black pupils were drawn by the Institute for Statistical Reseach (ISR) of the HSRC in the following manner: In the case of white schools in the Pretoria-Witwatersrand-Vereeniging (PWV) area the total number of high schools was divided into three strata according to language medium, namely Afrikaans-medium, English-medium and Afrikaans/English-medium schools. Twenty schools were then drawn at random and in proportion to the number of schools in each stratum (ten Afrikaans-medium, eight English-medium and two Afrikaans/English-medium schools). In the Cape Peninsula ten schools were drawn in a similar manner (two Afrikaans-medium, four English-medium and four Afrikaans/English-medium schools). In the case of Indian pupils 30 high schools were selected at random from the list of high schools in and around Durban.

In respect of the black pupils 30 schools were drawn at random from the list of high schools in the PWV area. Owing to the unrest situation the test programme could be carried out at only three of these schools, and with the co-operation of the Psychological Services of the Department of Education and Culture of KwaZulu the test programme was moved to Natal. The head of the Psychological Services was informed of the nature of the sample that was required, namely that the schools included in the sample had to be as close as possible to active economic centres and that schools should not be selected deliberately, for instance because they were the best schools. No further control could be exercised over the sample of black schools. The number of schools from the various areas in KwaZulu that were involved in the investigation was the following: eight schools from the Pietermaritzburg area, eight from Empangeni and nine from Newcastle.

The total sample is represented in Table 3.1.

TABLE 3.1

WHITE, INDIAN AND BLACK SAMPLES: NUMBER OF SCHOOLS AND PUPILS (STANDARD 7)

Area	White		Indian		Black	
	Schools	Pupils	Schools	Pupils	Schools	Pupils
Pretoria-Witwatersrand-Vereeniging	20	712			3	103
Cape Town	10	344				
Durban			30	1 063		
KwaZulu (Pietermaritzburg, Empangeni and Newcastle)					25	990
TOTAL	30	1 056	30	1 063	28	1 093

The ages of the pupils in the three samples are indicated in Table 3.2.

TABLE 3.2

MEAN ($\bar{X}$) AND STANDARD DEVIATION(S) OF THE AGE (IN YEARS) OF THE THREE POPULATION GROUPS AND DISTRIBUTION OF PUPILS BY AGE GROUP

Age (in years)	White		Indian		Black	
	N= 1 056 $\bar{X}$= 15,0 yrs s= 0,86 yrs		N= 1 063 $\bar{X}$= 15,0 yrs s= 0,99 yrs		N= 1 093 $\bar{X}$= 16,5 yr s= 1,67 yrs	
	N	%	N	%	N	%
13	8	0,8	8	0,8	60	5,5
14	545	51,7	564	53,1	127	11,6
15	398	37,7	376	35,4	248	22,7
16	77	7,3	88	8,3	289	26,4
17	22	2,1	24	2,3	173	15,8
18	-	-	-	-	122	11,2
19	-	-	-	-	32	2,9
20	-	-	-	-	42	3,8

Table 3.2 reveals that the black pupils were considerably older than the white and Indian pupils. For instance 89 % of the white and Indian pupils fell in the age group 14 years 0 months to 15 years 11 months, while only 34 % of the black pupils fell in this age group. In contrast only 9 % of the white and 10 % of the Indian pupils were older than 15 years 11 months, as opposed to 60 % of the black pupils.

3.2 MEASURING INSTRUMENTS

Four measuring instruments were used, namely the Junior Aptitude Tests (JAT) (which is the subject of the investigation), the Raven Standard Progressive Matrices, a reasoning test and a biographical questionnaire. The latter two instruments were constructed especially for the investigation. As this report is mainly concerned with the aptitude tests, the results of the reasoning test and the biographical questionnaire will be discussed in a separate report.

3.2.1 The Junior Aptitude Tests (JAT)

The Junior Aptitude Tests (subsequently called the JAT) are used on a large scale by white education departments in South Africa for subject and vocational guidance. Similar test batteries also exist for the other population groups in South Africa, but as these batteries are not identical to the JAT, the various groups cannot be compared directly with each other. In the present investigation an indication is given of the extent to which the JAT was suitable for other population groups as well, in other words the extent to which it was free of item bias and bias in predictive and construct validity.

In 1974, with a view to standardization, the JAT was administered to representative samples of white pupils in Standards 5, 6 and 7 and in 1977 to pupils in Standard 8. The JAT replaced the NB Aptitude Tests (Junior) that had been released in 1961 and were used until 1975.

According to the JAT's Manual (1980) the battery was based on the following definition of aptitude:

> *Aptitude is the potential that a person has that enables him, with a given amount of training and/or practice, to reach a certain level of ability.*
>
> *Aptitude, together with interest, attitude, motivation and other personality traits, will to a large extent determine the ultimate success of a person.* (pp. 3 and 5).

In the selection of the 10 tests of which the JAT consists the assumption was that a pupil's ability to
. reason on the basis of verbal and non-verbal material,
. store meaningful material as well as abstract symbols in his memory,
. make quick and accurate observations and
. visualize spatial configurations accurately

provide an indication of the extent to which he is able to complete an academic, technical and practical school course successfully.

The following is a brief description of the 10 tests.

Test 1: Classification

Each item consists of five drawings of familiar objects of which four have at least one common characteristic. The testee is expected to indicate the object that does not belong with the other four. The test therefore requires inductive reasoning with the aid of non-verbal material.

Test 2: Reasoning

The test consists of items that include number series, arithmetic problems, word analogies, vocabulary and general problems of a logical nature and it measures general verbal reasoning ability. Together with Test 1 this test also gives an indication of general intelligence.

Test 3: Number Ability

The test consists of the four main mathematical computations, namely addition, subtraction, division and multiplication and also includes knowledge of percentages. As the testee is expected to perform simple calculations quickly and accurately, the test consequently measures number ability.

Test 4: Synonyms

Each item comprises a given word and five others from which the testee has to select a synonym for the given word. In the first place the test measures vocabulary and in a wider sense verbal understanding as well.

Test 5: Comparison

At each item the testee has to select from five groups of symbols consisting of letters and numbers the one that corresponds exactly with a given group. The test therefore measures perceptual speed.

Test 6: Spatial 2-D

Each item consists of a drawing of a two-dimensional rectilinear figure and five similar drawings. The testee must choose from the set of five figures the one that will form a square with the given figure if they are joined correctly. The test measures two-dimensional spatial perception.

Test 7: Spatial 3-D

The test consists of 30 items that are divided into three groups. In the case of each item in the first group the testee has to indicate which drawing from five possibilities should be added to a given figure to form a cube. In respect of the second group the testee has to indicate which one of five three-dimensional figures is formed after the given two-dimensional figure has been folded/rolled up correctly. In the case of the last group the testee has to indicate which one of five three-dimensional figures will fit a given three-dimensional figure. The test measures three-dimensional spatial perception.

Test 8: Memory (Paragraph)

The test consists of paragraphs that the testee has to memorize. After having answered a few other tests, the testee has to answer questions on the content of these paragraphs. The test measures memory of meaningful material.

Test 9: Memory (Words and Symbols)

The test consists of two parts, namely word pairs and symbol pairs. Each pair consists of one familiar and one unfamiliar element. Immediately after the memorization period one member of each pair is given to the testee and he has to identify the other member from a selection of five possibilities. The test therefore measures the ability to form associations with meaningless material and to memorize these associations.

Test 10: Mechanical Insight

To measure mechanical ability a drawing is used to illustrate the operation of a mechanical apparatus or a mechanical or physical principle. One or more questions are asked about each drawing.

No results on the construct validity of the JAT's tests are mentioned in the manual. By combining some of the JAT's tests an estimated IQ score can be obtained, which produced a correlation of 0,8 with the NSAGT total IQ in the case of the standardization sample (white pupils only). Roos (1984) also found that the JAT can distinguish effectively between persons who undertake specific studies later at university.

The predictive validity of the JAT is based on correlations between the test scores and examination marks at school. In the case of Standard 7 pupils (standardization group) the correlations were in the region of 0,4, which can be regarded as reasonably satisfactory. The reliability of the tests for Standard 7 pupils (standardization group) was calculated with the aid of the Kuder-Richardson Formula 20 and varied between 0,69 (Classification) and 0,89 (Spatial 2-D). Although these coefficients were not particularly high, they were nevertheless satisfactory in the light of the heterogeneous nature of some of the tests.

Procedure regarding testing time

Owing to the fact that testing time can play such an important role in the test performances of the various population groups three testing times were used in the administration of the JAT. The testing times are indicated in Table 3.3. Time 2 in this table is the normal or standard testing time given in the Manual of the JAT; Time 1 is a little shorter and Time 3 is a little longer than the standard time. Although the manual does not mention how the testing time (Time 2 in Table 3.3) was determined, it can be assumed that the usual procedure for determining the testing time of a power test was followed in this case. This is the time required by 90 % of the testees to complete a test. With the exception of the JAT 5, JAT 8 and JAT 9 that are speed tests, the above method was probably followed in the case of all the tests of the JAT.

The lengths of Time 1 and Time 3 were determined arbitrarily. Time 1 is generally about three minutes shorter and Time 3 is about four minutes longer than the normal testing time. At the end of Time 1 the pupils were instructed to encircle on the answer sheet the number of the question with which they were busy at that stage, and this instruction was repeated at the end of Time 2. A testee who had completed a test within Time 1 therefore did not mark an item. Those who had completed the test in Time 2 marked one item, while those who had completed the test in Time 3 or were still busy by the end of Time 3 marked two items. Although this method does have shortcomings - for instance a testee could neglect to mark the item concerned and thus create an erroneous impression regarding the time it took to complete the test - the assumption is nevertheless that a reasonably accurate indication can be obtained for each of the groups with regard to the effect a longer testing time has on test performance.

TABLE 3.3
TESTING TIMES FOR THE JAT

JAT	Time 1	Time 2	Time 3
Test 1 : Classification	6 min.	9 min.	13 min.
Test 2 : Reasoning	9 min.	12 min.	16 min.
Test 3 : Number Ability	12 min.	15 min.	19 min.
Test 4 : Synonyms	5 min.	8 min.	12 min.
Test 5 : Comparison	5 min.	10 min.	12 min.
Test 6 : 2-D	9 min.	12 min.	16 min.
Test 7 : 3-D	9 min.	12 min.	16 min.
Test 8*: Memory (Paragraph)			
- Learning time		5 min.	
- Testing time		7 min.	
Test 9*: Memory (Words and Symbols)			
Strange words			
- Learning time		3 min.	
- Testing time		5 min.	
Symbols			
- Learning time		3 min.	
- Testing time		4 min.	
Test 10: Mechanical Insight	17 min.	20 min.	24 min.
NOTE: * It is not possible to use more than one testing time.			

3.2.2 The Raven Standard Progressive Matrices (SPM)

The Standard Progressive Matrices (SPM) consists of 60 matrix-type items that are grouped in five sets of 12 items each. Each set involves different principles according to which the patterns in the matrices vary and in each set the items become progressively more difficult. The test instructions are so easy that they can be explained with gestures. No time limit is set and all the testees are allowed to complete the test.

With regard to construct validity the SPM is generally seen as measuring fluid intelligence (McKenna 1984) and it provides a particularly pure measurement of Spearman's "g" factor (Jensen 1980). Some researchers estimate the g loading to be in the region of 0,80 (Raven *et al.* 1977). Factor analysis revealed that the SPM has high loadings on a factor that is common to most intelligence tests. In addition the SPM measures other specific factors such as inductive reasoning, spatial ability and perceptual accuracy to a lesser extent (Anastasi 1976).

The predictive validity of the SPM in respect of school performance is estimated at approximately 0,7 by Raven *et al.* (1977).

As far as the reliability of the SPM is concerned the coefficients vary between 0,6 and 0,9 (Anastasi 1976). The test-retest reliabilities with a period of a year between testings vary from 0,83 to 0,93 (Raven *et al.* 1977) and can be regarded as very high. High reliability coefficients

(K-R 20 $\geq$ 0,9) are also often found among different cultural groups (Carlson & Jensen 1981).

From the above it is clear that the SPM is a very suitable measuring instrument for use in cross-cultural research. In the present study the SPM was used in the following manner:

In the first place it served as an external criterion for determining the predictive validity of the JAT 1 (Classification). The SPM therefore plays the role that would have been played by examination marks if all the pupils had written the same school examinations.

In the second place the total scores of the SPM were used to divide the three population groups into comparable/homogeneous ability groups with a view to applying the Scheuneman and the full chi-square techniques for identifying biased items in the tests of the JAT (cf. Par. 7.2.1). Usually the total scores of the test concerned are used to form comparable ability groups. A disadvantage of this method is that if there are many biased items in the test, the total score is also biased. This problem can be solved by using another test that is less biased. In this regard, too, the SPM has proved its value (Shepard, Camilli & Averill 1981).

Lastly, in the application of the item response theory for identifying biased items in the JAT, the scores in the SPM are used as external criteria in estimating the ability levels (θ's) of the testees (cf. Par. 7.2.2).

3.3 PROCEDURES, METHODS AND TECHNIQUES

The investigation was conducted in conjunction with the Transvaal Education Department (TED), the Cape Education Department (CED), the Department of Education and Culture, Administration: House of Delegates, the Department of Education and Training and the Department of Education and Culture of KwaZulu. The tests were administered in August and September 1985 in the white and Indian schools, in November 1985 in the three schools of the Department of Education and Training and in June 1986 in the schools in KwaZulu. The testing was done by officials of the various education departments.

The thoroughness with which the ten tests of the JAT were investigated depended on the nature of the tests and the conceptual contribution such an analysis could make to the principles underlying the nature and essence of bias. For example all ten the tests were involved in analyses of the constructs measured by the JAT, but at item level three tests - the JAT 5 (Comparison) and the two memory tests (JAT 8 and JAT 9) - were not analysed further because performance in these cases was mainly a function of testing time. In contrast, special attention was paid to the JAT 1 (Classification) and the JAT 2 (Reasoning). These two tests are representative of tests for general intellectual ability and analyses here can throw more light on the effect of bias in cognitive tests in general.

The following methods and techniques are among those that were used to analyse the data:

. Item analysis. Item analysis was done in respect of Tests 1, 2, 3, 4, 6, 7 and 10 of the JAT. The item analysis programme concerned

provides information regarding the item's p values (in other words the ratio of the number of testees who have answered the item correctly to the number who have attempted to answer the item), the item discrimination values (point biserial correlation between the item and the test's total score) and the responses to the distractors (number of testees who have selected a specific distractor). These results appear in Appendix A.

- Factor analysis. Factor analysis by means of the principal factor analysis method was carried out on the intercorrelations of the ten tests of the JAT to study the similarity of constructs among the white, Indian and black pupils.
- Chi-square tests. Chi-square tests, namely those by Scheuneman and the full chi-square, were used to identify biased items in Tests 1, 2, 3, 4, 6, 7, and 10 of the JAT.
- Item response theory. The three-parameter logistical model (cf. Par. 7.2.2) was used in addition to the chi-square tests to identify biased items in the JAT 1 and JAT 2.

The above and other analysis techniques are discussed later in more detail together with the data concerned.

4.0 GENERAL FUNCTIONING OF THE JAT FOR INDIAN AND BLACK PUPILS

Before bias as such is discussed it is useful to take note of the general functioning of the JAT in respect of the various population groups. In this regard attention is given to the test means and standard deviations, the normality of the score distributions and the effect of testing time on test performance.

4.1 MEANS AND STANDARD DEVIATIONS

The test means and standard deviations that were obtained with the standard testing time (Time 2), as well as the significance of the differences in means (calculated with the aid of Hotelling's T^2 and *post hoc* t tests) appear in Table 4.1. This table reveals that all the differences are statistically significant ($p < 0{,}0001$) and that black pupils in particular performed less well than their white peers. The actual size of the differences can be seen in Table 4.2, where the difference in the means for each test is expressed in terms of the standard deviation concerned of the white pupils (the differences between white/Indian and white/black can therefore be compared directly).

TABLE 4.1
THE MEANS ($\bar{X}$) AND STANDARD DEVIATIONS (S) OF THE JAT FOR WHITE, INDIAN AND BLACK PUPILS AND THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE GROUPS (HOTELLING T^2 AND t TESTS)

Tests	Maximum score	White		Indian		Black		Differences between White/Indian		Differences between White/ Black	
		$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	S	df	t	df	t
JAT 1: Classification	30	23,1	3,5	19,1	4,1	13,4	4,3	2 117	24,1*	2 147	56,8*
JAT 2: Reasoning	30	18,3	4,8	13,7	4,7	6,5	2,6	2 117	22,3*	2 147	70,6*
JAT 3: Number Ability	30	16,7	4,9	15,6	4,8	9,0	3,6	2 117	5,1*	2 147	41,2*
JAT 4: Synonyms**	30	19,9	4,9	14,5	5,6	6,1	2,7	2 117	23,5*	2 147	80,4*
JAT 5: Comparison	30	24,6	3,9	20,6	5,7	16,5	5,5	2 117	18,7*	2 147	39,0*
JAT 6: 2-D	40	27,8	7,2	17,4	6,9	13,6	5,9	2 117	33,9*	2 147	49,9*
JAT 7: 3-D	30	22,0	5,5	14,5	6,4	10,3	4,9	2 117	28,8*	2 147	51,7*
JAT 8: Memory (Paragraph)	25	17,4	4,6	13,6	4,3	8,2	3,8	2 117	20,0*	2 147	50,6*
JAT 9: Memory (Symbols)	30	22,7	5,5	19,5	5,6	16,7	5,9	2 117	13,2*	2 147	24,4*
JAT 10: Mechanical Insight	42	19,7	5,5	12,2	4,0	7,9	3,0	2 117	35,5*	2 147	61,7*
	N	1 056		1 063		1 093		T^2=2 194,5*		T^2=9 359,4*	

NOTE: *p<0,0001
** ALTHOUGH THE AFRIKAANS AND ENGLISH VERSIONS OF THE TEST DIFFER, SIMILAR MEANS AND VARIANCES WERE FOUND AND THE DATA FOR THE TWO GROUPS COULD THEREFORE BE POOLED.

TABLE 4.2

DIFFERENCES BETWEEN THE TEST MEANS OF WHITE AND INDIAN AND WHITE AND BLACK PUPILS IN TERMS OF THE STANDARD DEVIATION OF THE WHITES

TEST	White/Indian	White/Black
JAT 1: Classification	1,1	2,8
JAT 2: Reasoning	1,0	2,5
JAT 3: Number Ability	0,2	1,6
JAT 4: Synonyms	1,1	2,8
JAT 5: Comparison	1,0	2,1
JAT 6: 2-D	1,4	2,0
JAT 7: 3-D	1,4	2,1
JAT 8: Memory (Paragraph)	0,8	2,0
JAT 9: Memory (Symbols)	0,6	1,1
JAT 10: Mechanical Insight	1,4	2,1

From Table 4,2 it is clear that the performance of black pupils was in some cases (JAT 1, JAT 2 and JAT 4) two-and-a-half to almost three standard deviation units lower than that of white pupils. However it should be kept in mind that the results were based on the standard testing time of the battery. The extent to which testing time played a role in test performance is discussed in Par. 4.3.

As far as Indian pupils are concerned it seemed that, in contrast with the blacks, their performance was approximately only one standard deviation lower than that of the white pupils. With regard to the JAT 3 (Number Ability) there was almost no difference in the performance of the two groups.

4.2 NORMALITY OF THE SCORE DISTRIBUTIONS

Data regarding the normality of the score distributions in respect of skewness and kurtosis are given in Table 4.3. As far as skewness is concerned it seems that few of the distributions were normal: the distributions of the white pupils were mainly negatively skewed, those of the Indian pupils mainly positively skewed, while the distributions of the black pupils were, with one exception, all positively skewed (the tests were therefore relatively easy for whites, more difficult for Indians and very difficult for blacks).

In the case of kurtosis (the peakedness of the curve) normal distributions were more common among the three population groups.

As far as the reliabilities (internal consistency, calculated with the Kuder-Richardson Formula 21) are concerned Table 4.3 indicates that there are relatively large differences between the white and Indian pupils on the one hand and the black pupils on the other, particularly in respect of the JAT 2, the JAT 4 and the JAT 10 (reliability is discussed further in Par. 6).

TABLE 4.3
SKEWNESS (SK), KURTOSIS (KU) AND RELIABILITY (K-R21) OF THE TEST SCORES OF THE JAT (STANDARD TESTING TIME) FOR WHITE, INDIAN AND BLACK PUPILS

Tests	White			Indian			Black		
	Sk	Ku	K-R21	Sk	Ku	K-R21	Sk	Ku	K-R21
JAT 1: Classification	-0,78	4,50	0,59	-0,28	2,85*	0,60	0,11*	2,81*	0,63
JAT 2: Reasoning	0,15	2,64	0,72	0,41	2,88*	0,68	0,46	3,38	0,27
JAT 3: Number Ability	0,14	2,80*	0,72	0,11*	2,59	0,70	0,43	3,31	0,54
JAT 4: Synonyms	-0,31	2,95*	0,75	0,33	2,55	0,79	0,44	3,24*	0,35
JAT 5: Comparison	-1,29	6,61	1)	-1,00	4,42	1)	-0,27	3,29	1)
JAT 6: 2-D	-0,33	2,37	0,86	0,49	2,75*	0,82	0,73	3,69	0,77
JAT 7: 3-D	-0,78	3,03*	0,84	0,24	2,21	0,85	0,75	3,28	0,75
JAT 8: Memory (Paragraph)	-0,27	2,30	0,78	0,15	2,66	0,69	0,84	4,09	0,65
JAT 9: Memory (Symbols)	-0,53	2,48	0,85	-0,07*	2,34	0,81	0,23	2,51	0,81
JAT 10: Mechanical Insight	0,27	3,10*	0,67	0,33	2,77*	0,48	0,41	3,14*	0,29

NOTE: ' K-R formulas not applicable to tests where speed plays a major role.
* Normality of this aspect of the distribution can be accepted on the 10 % confidence level (two-tailed test).

To summarize it can be said that, as far as means, score distributions and reliabilities are concerned, the JAT in general does not function well for black pupils.

4.3 TESTING TIME

In Par. 3 it was stated that three testing times had been used in administering the battery. Time 2 was the normal testing time as prescribed in the manual of the JAT, Time 1 was a bit shorter and Time 3 was a bit longer than the normal testing time. Naturally only one testing time could be used in the case of the memory tests and the standard testing time seemed to be the best choice.

The proportion of testees who had reached a certain test item within a specific period of time is represented in Table 4.4 (the results of only five tests are given as the results of the others are similar). As expected the proportion of testees who had reached a certain item depended on the testing time. However there was a marked difference in the proportions of, in particular, the white and black testees who had reached the last third of the test items: 89 % of the white as opposed to 21 % of the black testees had reached Item 21 of the JAT 1 (Classification) in Time 1. It is also significant that, in a purely perceptual speed test such as the JAT 5 (Comparison) in which higher-order cognitive functioning such as reasoning does not play an important role, the difference in work pace is evident to the same extent as in other tests.

TABLE 4.4

PROPORTION OF THE TESTEES WHO HAD REACHED A TEST ITEM IN A CERTAIN TESTING TIME (T)

NOTE: (a) JAT 1: Classification									
Item number	White N=1 020)*			Indian (N=1 063)			Black (N=1 093)		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
1	0,99	0,99	0,99	0,99	0,99	0,99	0,98	0,98	0,98
5	0,99	0,99	0,99	0,99	0,99	0,99	0,97	0,97	0,97
10	0,99	0,99	0,99	0,99	0,99	0,99	0,88	0,96	0,97
15	0,99	0,99	0,99	0,90	0,98	0,99	0,53	0,88	0,97
18	0,97	0,99	0,99	0,75	0,98	0,99	0,36	0,76	0,97
21	0,89	0,99	0,99	0,51	0,94	0,99	0,21	0,61	0,94
24	0,77	0,98	0,99	0,30	0,85	0,98	0,12	0,48	0,91
27	0,59	0,95	0,99	0,13	0,69	0,97	0,05	0,32	0,83
30	0,38	0,88	0,99	0,05	0,52	0,97	0,02	0,25	0,79
NOTE: (b) JAT 2: Reasoning									
1	0,99	0,99	0,99	0,99	0,99	0,99	0,97	0,99	0,99
5	0,99	0,99	0,99	0,98	0,99	0,99	0,97	0,99	0,99
10	0,99	0,99	0,99	0,98	0,99	0,99	0,91	0,98	0,99
15	0,99	0,99	0,99	0,93	0,98	0,99	0,47	0,87	0,95
18	0,98	0,99	0,99	0,80	0,97	0,99	0,29	0,74	0,94
21	0,91	0,98	0,98	0,55	0,90	0,97	0,16	0,54	0,90
24	0,65	0,91	0,95	0,22	0,65	0,88	0,06	0,31	0,81
27	0,36	0,81	0,95	0,08	0,43	0,83	0,03	0,22	0,71
30	0,16	0,60	0,95	0,03	0,27	0,82	0,02	0,15	0,65
NOTE: (c) JAT 3: Number Ability									
1	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99
5	0,99	0,99	0,99	0,99	0,99	0,99	0,98	0,98	0,98
10	0,98	0,98	0,98	0,97	0,98	0,98	0,79	0,94	0,96
15	0,93	0,96	0,96	0,87	0,96	0,97	0,31	0,64	0,86
18	0,83	0,95	0,96	0,64	0,90	0,97	0,13	0,44	0,80
21	0,57	0,87	0,95	0,34	0,73	0,91	0,05	0,28	0,69
24	0,33	0,71	0,90	0,15	0,51	0,83	0,02	0,15	0,58
27	0,15	0,48	0,79	0,05	0,28	0,74	0,01	0,10	0,52
30	0,06	0,28	0,68	0,01	0,15	0,65	0,01	0,08	0,46

TABLE 4.4 (continued)

Item number	White N=1 020)*			Indian (N=1 063)			Black (N=1 093)		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
Note: (d) JAT 5: Comparison									
1	1,00	1,00	1,00	1,00	1,00	1,00	0,99	0,99	0,99
5	1,00	1,00	1,00	0,99	1,00	1,00	0,98	0,99	0,99
10	0,98	1,00	1,00	0,89	0,99	1,00	0,73	0,98	0,99
15	0,75	1,00	1,00	0,54	0,97	0,99	0,30	0,93	0,99
18	0,44	0,99	1,00	0,25	0,94	0,98	0,14	0,81	0,97
21	0,19	0,98	0,99	0,11	0,88	0,96	0,07	0,63	0,92
24	0,08	0,94	0,98	0,05	0,72	0,92	0,03	0,44	0,82
27	0,04	0,81	0,96	0,02	0,53	0,84	0,02	0,27	0,73
30	0,03	0,64	0,92	0,02	0,37	0,79	0,01	0,19	0,67
Note: (e) JAT 6: 2-D									
1	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99
5	0,99	0,99	0,99	0,98	0,98	0,99	0,99	0,99	0,99
10	0,99	0,99	0,99	0,98	0,98	0,99	0,97	0,98	0,99
15	0,99	0,99	0,99	0,93	0,98	0,99	0,85	0,97	0,99
20	0,96	0,99	0,99	0,74	0,95	0,98	0,62	0,91	0,98
25	0,83	0,97	0,99	0,44	0,81	0,96	0,33	0,74	0,94
30	0,60	0,90	0,99	0,20	0,61	0,91	0,14	0,53	0,88
35	0,35	0,76	0,96	0,07	0,38	0,81	0,06	0,33	0,80
40	0,17	0,60	0,94	0,03	0,22	0,77	0,02	0,21	0,73
Note: * N<1 056 because the data of one school were disregarded since only one time limit (Time 2) had been used.									

Figure 4.1 is a graphical representation of the proportion of testees who reached the last item of each of the five tests contained in Table 4.4. As can be seen a larger proportion of white than Indian or black pupils reached the last item. This applied throughout in respect of all the tests and for each of the three testing times.

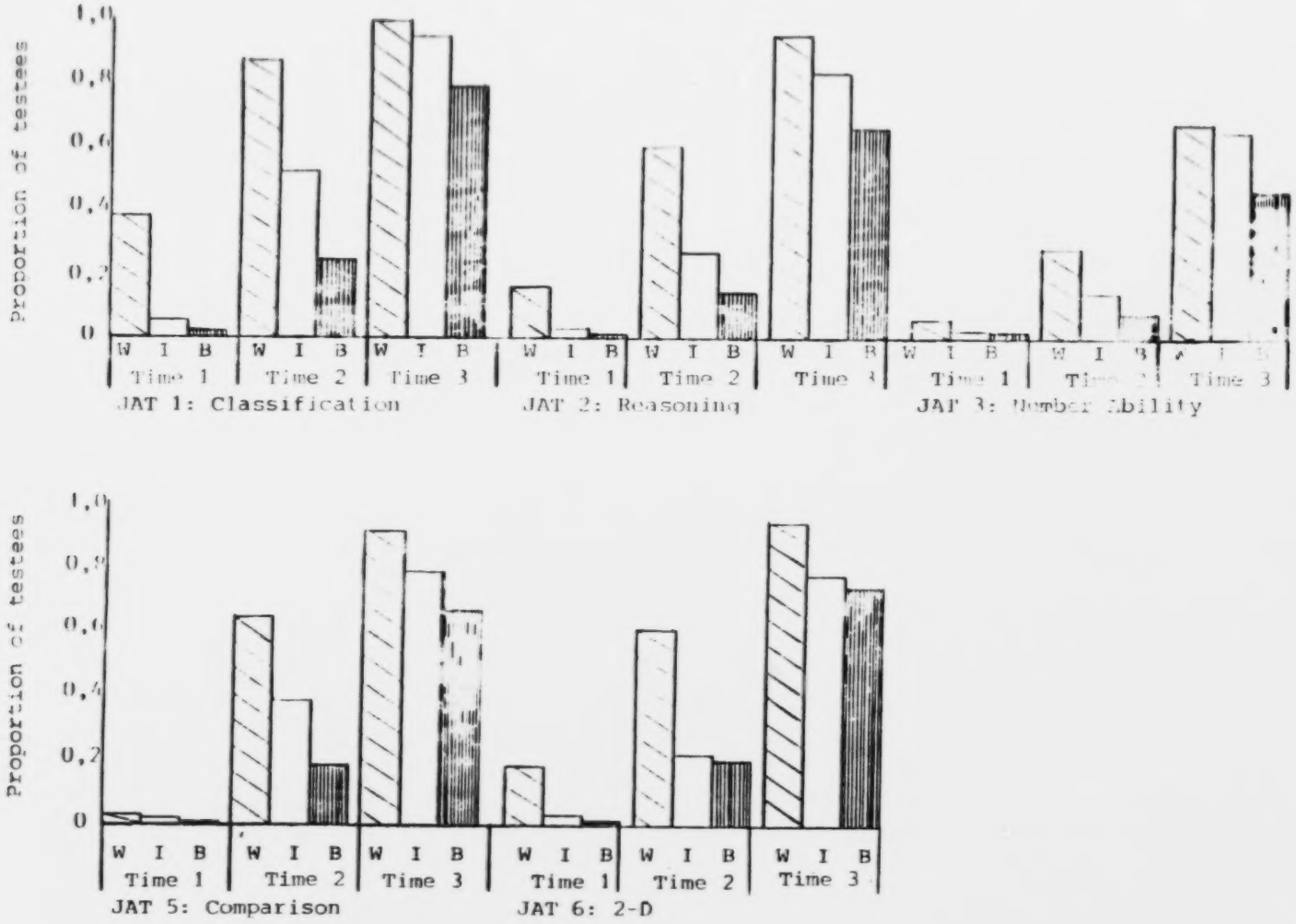


FIGURE 4.1 REPRESENTATION OF THE PROPORTION OF THE WHITE (W), INDIAN (I) AND BLACK (B) PUPILS WHO HAD REACHED THE LAST ITEM IN A TEST IN A CERTAIN TESTING TIME

The above findings indicate that the black pupils processed information at a slower rate than the white pupils.

The question is now whether a longer testing time implies an improved mean test performance. Table 4.5 reveals that the average improvement in test performance from Time 1 to Time 3 for white, Indian and black pupils was 3,1, 4,3 and 3,8 points respectively (the JAT 5 was not included in the calculation as it is a speed test and time therefore plays a more important role than in any of the other tests). Since the increase for the black pupils does not differ significantly from those for the two other groups (the Bonferroni confidence intervals overlap), the conclusion is that the effect of longer testing time is not greater among the black pupils than among the two other population groups.

The above results concern the population groups as a whole. To study the effect of testing time further, three non-overlapping groups in each population group were formed on the basis of the test scores that had been obtained in the case of a certain testing time. Group A consisted of those testees who had completed a specific test in Time 1, Group B of those who had completed the test in Time 2, and Group C of all the remaining testees, in other words those who had completed the test in Time 3 or had still been busy at the end of Time 3. Table 4.6 contains the test means of these three groups for five tests of the JAT in respect of the individual population groups. This table reveals that, in the case of the white testees, Group A not only worked the fastest but in general also obtained the highest scores. In the case of the Indian testees Group A did work quickly, but they also made more mistakes than Groups B and C, who worked more slowly. In the case of the black testees the performance of the various groups depended on the type of test being answered: Group A performed the best in the JAT 2 and JAT 5, Group B in the JAT 1 and Group C in the JAT 3 and JAT 6. With the longer testing time available the latter group, which included about 76 % to 89 % of the total sample of black testees, performed better than Groups A and B (who required less time to complete the tests) in only two of the five tests.

From Table 4.6 it seems as if there could be an interaction between the various population groups and the time groups (A, B and C). To investigate this matter a two-way analysis of variance was made in respect of three of the five tests, with population group and time group as independent variables and test score as the dependent variable. A three (population groups) X three (time groups) type of design was obtained. Table 4.7 gives a summary of the analysis of variance of the JAT 1 the JAT 2 and the JAT 3. As expected some tests revealed a significant ($p < 0,01$) interaction between population and time group. However the contribution of this source to the total variance was extremely small - in most cases it was less than 1 %. As far as the two main effects are concerned only the population group was significant throughout ($p < 0,001$) and this also explains a large percentage of the variance in the scores.

To summarize it can be stated that testing time as a possible cause of test differences between groups does not play such an important role as is often surmised.

TABLE 4.5
TEST MEANS OF THE WHITE, INDIAN AND BLACK PUPILS WITH REGARD TO A CERTAIN TESTING TIME(T)

Tests	White (N=1 020)							Indian (N=1 063)							Black (N=1 093)						
	T1		T2		T3		Difference	T1		T2		T3		Difference	T1		T2		T3		Difference
	$\bar{X}1$	S	$\bar{X}2$	S	$\bar{X}3$	S	$\bar{X}3-\bar{X}1$	$\bar{X}1$	S	$\bar{X}2$	S	$\bar{X}3$	S	$\bar{X}3-\bar{X}1$	$\bar{X}1$	S	$\bar{X}2$	S	$\bar{X}3$	S	$\bar{X}3-\bar{X}1$
JAT 1: Classification	20,9	4,3	23,1	3,5	23,3	3,4	2,4	15,1	4,1	19,1	4,0	20,3	3,8	5,2	9,9	4,0	13,4	4,3	16,1	4,0	6,2
JAT 2: Reasoning	16,9	4,5	18,3	4,9	18,9	5,0	2,0	12,1	4,1	13,7	4,7	14,7	5,0	2,6	5,3	2,4	6,5	2,6	7,5	2,7	2,2
JAT 3: Number Ability	14,9	4,7	16,7	5,0	18,0	5,1	3,1	13,3	4,3	15,6	4,8	17,4	5,1	4,1	7,2	3,2	9,0	3,6	10,9	4,1	3,7
JAT 4: Synonyms	17,3	5,1	19,9	4,9	20,1	4,9	2,8	11,1	5,0	14,5	5,6	15,6	5,5	4,5	4,4	2,4	6,1	2,7	7,2	2,7	2,8
JAT 5: Comparison	14,9	4,1	24,6	3,9	25,7	3,5	10,8	12,0	4,6	20,6	5,7	23,0	5,5	11,0	9,4	4,4	16,5	5,5	20,1	5,5	10,7
JAT 6: 2-D	23,8	7,2	27,8	7,2	29,4	6,8	5,6	13,8	5,5	17,4	6,9	20,2	7,8	6,4	10,5	4,9	13,6	5,9	16,2	6,8	5,7
JAT 7: 3 D	20,7	5,8	22,0	5,6	22,2	5,5	1,5	12,1	5,5	14,5	6,4	15,4	6,7	3,3	8,5	4,2	10,3	5,0	11,2	5,3	2,7
JAT 10: Mechanical Insight	17,3	5,3	19,7	5,5	21,3	5,6	4,0	10,3	3,5	12,2	4,0	14,3	4,6	4,0	6,3	2,7	7,9	3,0	9,7	3,3	3,4
Average increase in test means (JAT 5 excluded):							3,1*							4,3*							3,8*

NOTE: * A 95 % Bonferroni confidence interval for $\bar{X}3 - \bar{X}1$ is as follows:
white : 2,74; 3,46
Indian : 3,94; 4,66
black : 3,44; 4,16

TABLE 4.6
MEAN TEST SCORES ON TESTS 1, 2, 3, 5 AND 6 OF THE JAT FOR TESTEES WHO COMPLETED A TEST IN TIME 1 (GROUP A), TIME 2 (GROUP B) AND TIME 3 (GROUP C*)

Tests	White (N=1 020)						Indian (N=1 063)						Black (N=1 093)					
	Group A		Group B		Group C		Group A		Group B		Group C		Group A		Group B		Group C	
	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N
JAT 1	23,6	341	23,3	547	22,5	132	19,4	49	20,5	441	20,2	573	16,4	31	16,6	227	16,0	832
JAT 2	19,9	135	18,8	456	18,6	429	12,2	27	14,8	233	14,8	802	7,9	25	7,6	132	7,4	932
JAT 3	18,4	53	18,1	213	17,9	754	12,1	14	17,3	138	17,5	911	9,8	16	10,6	92	11,0	975
JAT 5	25,0	30	26,0	599	25,3	391	22,1	19	23,7	336	22,8	707	21,6	17	20,9	195	19,9	875
JAT 6	30,3	170	29,9	430	28,6	420	14,3	32	18,8	202	20,8	829	13,2	32	15,4	200	16,5	859

NOTE: * Includes pupils who were still busy at the end of Time 3

BEST COPY AVAILABLE

TABLE 4.7
SUMMARY OF ANALYSIS OF VARIANCE OF THE SCORES ON TESTS 1, 2 AND 3 OF THE JAT WITH REGARD TO POPULATION AND TIME GROUP

Source of variation	Sum of squares	df	Mean square	F	Level of significance	η^2(x100)'
(a) JAT 1: Classification						
Population group (B)	27 528,7	2	13 764,4	1 002,3	p<0,0001	38,7
Time group (T)	146,9	2	73,5	5,4	p<0,01	0,2
B x T	124,8	4	31,2	2,3	p<0,06*	0,2
Fault	43 327,7	3 155	13,7			60,9
Total	71 128,1	3 163				100,0
(b) JAT 2: Reasoning						
Population group (B)	70 178,7	2	35 089,4	1 881,2	p<0,0001	54,3
Time group (T)	52,1	2	26,0	1,4	p<0,25*	0,04
B x T	331,1	4	82,8	4,4	p<0,001	0,26
Fault	58 793,9	3 152	18,6			45,4
Total	129 355,8	3 160				100,0
(c) JAT 3: Number Ability						
Population group (B)	31 533,6	2	15 766,8	689,4	p<0,0001	30,6
Time group (T)	90,5	2	45,3	2,0	p<0,13*	0,09
B x T	382,6	4	95,7	4,2	p<0,01	0,4
Fault	71 061,9	3 107	22,9			68,9
Total	103 068,6	3 115				99,9

NOTE: * Not significant
'Eta squared is the sum of squares for each source divided by the total sum of squares; multiplied by 100, it is the percentage of the total variation in the data accounted for by each source.

5.0 BIAS IN THE PREDICTIVE VALIDITY OF THE JAT

In its simplest form test validity means that the test measures what it is supposed to measure (Burrill & Wilson 1980); it also refers to the applicability of the conclusions that are based on the test scores (Boehm 1978). Although several forms of validity are distinguished (Burrill 1981; Messick 1980), only criterion-related or predictive validity will be discussed here. This again involves two types of validity, namely concurrent validity (test and criterion data were obtained at the same time) and predictive validity (test and criterion data were obtained at different times). However these two forms of criterion-related validity are calculated in exactly the same way.

5.1 Differential and single-group validity

Differential validity occurs when there is a significant difference in the correlations between the test and criterion scores of different groups (Bond 1981).

The evaluation of bias in predictive validity in terms of the regression definition by Cleary (1968) takes the form of $\hat{Y}_i = aX_i + b$, where $\hat{Y}_i$ is the predicted criterion score, a is the regression coefficient, X_i is the score in the predictor (test) and b is a constant. If the equation is represented graphically in the form of a regression line, a represents the slope and b the Y intercept of the line. Differential predictive validity occurs when the slopes of the regression lines for two or more groups differ and the same regression line is used for all the groups. Bias in this case lies in the fact that, according to the definition of bias in predictive validity, the errors in the prediction should be independent of group membership - which is not the case if the slopes of the regression lines of the various groups differ. In other words the expected errors of estimate are not zero (Anastasi 1988; Cole 1981; Petersen 1980; Shepard 1982; Van der Flier & Drenth 1980).

Bias in respect of differential predictive validity also means that the correlations between the test and criterion scores (validity coefficient) for various groups differ significantly from each other and that the varying degree of validity is a function of group membership (Shepard, Camilli & Averill 1981). The differential validity hypothesis is therefore that certain tests are more valid for the majority than the minority groups, in other words $\rho_A < \rho_B$ (Hunter & Schmidt 1978), but that the coefficients are not necessarily zero for either of the two groups (Schmidt, Pearlman & Hunter 1980).

According to the single-group validity hypothesis the same test would be valid for one group but not for another, in other words the true population validity is zero for the one group but not necessarily for the other group as well (Hunter & Schmidt 1978). Single-group validity is therefore the result of the predictive validity coefficients for only one of the groups involved being greater than zero, in other words $0 = \rho_A < \rho_B$ (Thorndike & Hagen 1969; Weiss & Davidson 1981). In its most extreme form this hypothesis of test bias maintains that, in the case of whites and blacks,

> ...*black culture is so alien to white culture that a test might be completely meaningless to blacks* (Hunter, Schmidt & Rauschenberger 1984, p45).

The Raven Standard Progressive Matrices (SPM) was used as the criterion in the study of the differential and single-group validity of the JAT. Owing to the fact that only one of the JAT's tests is conceptually similar to the criterion in the sense that both measure non-verbal reasoning ability, bias in predictive validity was investigated only in respect of the JAT 1 (Classification). The product moment correlation between the scores of the JAT 1 (that were obtained with the standard testing time) and the SPM was calculated. The "value" of these coefficients was determined by means of confidence intervals.

The possible size of these coefficients if the unreliability of the predictor and the criterion had not led to attenuation was determined with the aid of the following formula:

$k^{r}xy = {}^{r}xy / \sqrt{{}^{r}xx\,{}^{r}yy}$, where $k^{r}xy$ is the corrected correlation, ${}^{r}xy$ is the correlation between the two variables and ${}^{r}xx$ and ${}^{r}yy$ are the reliability of the two variables. The results appear in Table 5.1.

From Table 5.1 it is clear that the correlations of each of the population groups differ significantly from zero and single-group validity, in other words the case where the correlation of one group differs significantly from zero but not that of the other group, is therefore not present.

As far as differential validity is concerned it is evident from both the z value and the overlapping confidence intervals that there are no significant differences between the correlations of the black and white testees. However a significant difference can be found between the correlations of the white and the Indian pupils and differential validity could be present (however owing to the fact that the two reliability intervals overlap to a small extent, this "possibility" is not very strong).

The above finding, namely that differential and single-group validity are mainly absent as far as the JAT1 is concerned, corresponds with most overseas findings in this regard. Jensen even calls it a ...*non-existent phenomenon*... (1980, p515). Hunter, Schmidt and Hunter (1979) investigated 3 validity studies and came to the conclusion that perhaps true differential validity does not exist. According to these writers the fact that they could not find any proof for the phenomenon of differential validity provides strong support against a hypothesis of test bias that is based on the assumption ...*that the meaning of test content differs by race* (p. 733). In contrast to the above writers, Katzell and Dyer (1978) maintain that it is too soon to reject the matter of differential validity completely. The fact that differential and single-group validity are seldom encountered or are difficult to identify indicates to Flaugher (1978) that these sources of bias are not very important in relation to others.

TABLE 5.1

PREDICTIVE VALIDITY: CORRELATION COEFFICIENTS (r_{xy}), SIGNIFICANCE OF THE DIFFERENCES BETWEEN POPULATION GROUPS (Z), CONFIDENCE INTERVALS AND CORRECTION FOR ATTENUATION ($_kr_{xy}$) FOR THE JAT 1 AND THE SPM

Type of calculation	White (N=1 056)	Indian (N=1 063)	Black (N=1 093)	Differences between:	
				White/Indian	White/Black
				z	z
Correlation: r_{xy} Interval for r_{xy} Corrected correlation: $_kr_{xy}$	0,284** 0,21≤r≤0,35 0,388	0,418** 0,35≤r≤0,49 0,548	0,314** 0,24≤r≤0,38 0,408	4,0**	0,697*
NOTE: ** p<0,01 * Not significant					

5.2 THE SLOPE, INTERCEPT AND STANDARD ERROR OF ESTIMATE OF THE REGRESSION LINE

Cleary (1968) and Cleary and Hilton (1968) were among the first to try and define test bias in terms of predictive validity:

> *A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup* (Cleary 1968, p115).

Therefore in terms of this definition a test is biased if the criterion score, which was predicted with the aid of the common regression line, is consistently too high or too low for the members of the subgroup. In accordance with Cleary's definition many researchers (such as Evans & Reilly 1972; Goldman & Hewitt 1975) agree that a test is unbiased if the regression lines of the groups concerned are identical. Under these circumstances group membership in terms of race or sex plays no role (Humphreys & Taber 1973).

Jensen's (1980) definition of bias in predictive validity is more comprehensive than Cleary's and all the most important elements that play a role in bias are mentioned in a statistical manner:

> *A test with perfect reliability is a biased predictor if there is a statistically significant difference between the major and minor groups in the slopes b_{yx}, or in te intercepts k, or in the standard error of estimates $SE\hat{y}$ of the regression lines of the two groups. Conversely, an unbiased test with perfect reliability is one for which the major and minor groups do not differ significantly in b_{yx}, k or $SE\hat{y}$* (p. 379).

From this definition it is clear that the emphasis falls on bias in the predictive validity of the test and not on some inherent characteristic of the test itself. If homogeneity of regression (the Cleary model) is not present, the following four conditions can occur:

1. The slopes (regression coefficients) differ.
2. The intercepts differ.
3. The slopes and intercepts differ.
4. The standard errors of estimate differ.

Bias owing to the different slopes of the regression lines of two groups leads to an underprediction of the criterion score of the group with the higher mean criterion score when a common regression line is used.

Bias owing to different intercepts (the intercept of a regression line refers to the point at which the regression line cuts the axes) while the slopes are the same leads to a systematic underprediction or overprediction of the criterion score by the test. Although a test may have the same validity coefficients for two groups, it could still be subject to intercept bias.

If both the slopes and intercepts of the groups differ, the amount of bias in the prediction can vary owing to the use of a common regression line, and the direction of the bias can even be reversed.

Bias in respect of the standard error of estimate: The standard error of estimate is an index of the number of errors that are made in respect of the prediction of the criterion, in other words the scatter of the obtained criterion scores around the predicted criterion scores. The exact meaning of $SE_{\hat{Y}}$ can, according to Jensen (1980, p.378), be represented accurately only in its mathematical form, namely $SE_{\hat{Y}}=\sqrt{\Sigma(Y-\hat{Y})^2/N-2}$, where Y is the obtained criterion scores, $\hat{Y}$ is the predicted criterion scores and N is the number of cases.

As far as bias regarding the slope, intercept and standard error of estimate of the regression line in respect of the JAT is concerned, the regression equation (which describes the relation between the scores in the JAT 1 (Classification) and the SPM) of the white testees was compared with those of the Indian and the black testees. This was done with the aid of a method that is described by Potthoff (Reynolds 1982, p.220-221). The results appear in Table 5.2.

TABLE 5.2

JAT 1 AND SPM: SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE POPULATION GROUPS IN THE SLOPES, INTERCEPTS AND STANDARD ERRORS OF ESTIMATE OF THE REGRESSION LINES

Groups compared	Slope and intercept (F1)	Slope (F2)	Intercept (F3)	Standard error of estimate
White/Indian	10,3**	18,6**	2,1[a]	1,51**
White/black	190,0**	8,0**	370,9**	3,11**

Note: ** $p<0,01$
[a] Not significant

*F*1 in Table 5.2 is a test statistic for the simultaneous testing of the hypothesis that the regression coefficients (slope) and the intercept are constant for groups, and it refers to the *F*-distribution with 2g-2 and N-2g degrees of freedom where g is the number of groups and N the number of testees. As *F*1 is statistically significant with regard to the comparison of both white with Indian and white with black, the presence of bias can be assumed. To determine whether the bias is located in the regression coefficient, in other words in the slope of the regression line, or in the intercept (or in both), *F*2 and *F*3 are calculated subsequently (if *F*1 is not significant the slope and intercept should not be analysed separately). *F*2, with g-1 and N-2g degrees of freedom, indicates the constancy of the regression coefficients for the groups. Table 5.2 reveals that in this case too there are significant differences between white and Indian as well as between white and black testees. *F*3, with g-1 and N-g-1 degree of freedom, tests the equality of the intercepts of the regression lines of the three groups. According to Table 5.2 there is no significant difference between the intercepts of the regression lines of white and Indian pupils, but the difference between those of white and black pupils is significant.

Before the conclusion can be made that there is no bias in predictive validity, it should be indicated that the standard errors of estimate of the groups are the same. The standard error of estimate of the criterion (SPM) is calculated with the aid of an equation provided by Reynolds (1982, p.223). Table 5.2 reveals that the standard errors of estimate in respect of both the comparison between white and Indian and between white and black testees are significant.

To summarize it can be stated that, as far as the slope, intercept and standard error of estimate are concerned, bias was found in respect of all three matters in the case of black pupils and in respect of two of the three matters in the case of Indian pupils (only the intercept was not biased).

5.3 THE USE OF A COMMON REGRESSION LINE

The question is now to what extent the performance of Indian and black testees in the criterion is overpredicted or underpredicted if the regression equation for the white testees is used. The regression equation for the prediction of a criterion score on the basis of a test score is the following: $y=ax+b$. In the case of the white testees the regression equation for predicting the mean performance in the SPM on the basis of the mean performance in the JAT 1 is the following:
$\hat{Y}_{SPM} = 0{,}511\bar{x}_{JAT\ 1} + 33{,}47$.

The results that were obtained by using this regression equation for the Indian and the black testees as well appear in Table 5.3.

TABLE 5.3

DEVIATION (D) IN STANDARD DEVIATION UNITS OF PREDICTED MEAN ($\hat{\bar{Y}}$) FROM OBTAINED MEAN ($\bar{Y}$) ON THE SPM FOR INDIAN AND BLACK PUPILS

Population group	Obtained mean on criterion (SPM)		Predicted mean on criterion (SPM)	Discrepancy	
	$\bar{Y}$	s	$\hat{\bar{Y}}$	D[1]	Dc[2]
Indian	42,0	8,2	43,2	0,15	0,01
Black	27,6	10,7	40,3	1,2	0,9

NOTE: [1] $D=(\hat{\bar{Y}}-\bar{Y})/\sigma$ is an index of the amount of bias in predictive validity; positive values of D indicate overprediction of the criterion score.
[2] Dc is the discrepancy after correction for attenuation (due to the unreliability of the predictor, i.e. JAT 1)

According to Table 5.3 the use of the white testees' regression equation leads to a considerable overprediction (1,2σ) of the mean criterion score for black pupils and even if provision were to be made for attenuation in the correlation coefficient of the predictor and the criterion owing to the unreliability of the predictor, the overprediction still remains almost one standard deviation unit (0,9σ). Bias in predictive validity is therefore to the advantage of black pupils. This finding

corresponds with the most common findings in the USA (cf. Jensen 1980). In the case of the Indian testees there was almost no difference between the actual and estimated criterion means.

According to Jensen (1980), criterion scores for blacks are overpredicted by a common regression equation because the intercept of the regression line of whites is usually higher than that of blacks. According to Sung and Dawis (1981) the finding that, in cases where the regressions of whites and blacks differ, the difference can usually be attributed to differences in intercepts and not in slopes, supports the assertion that differences in the mean test performance of groups do not serve as an indication of bias.

6.0 BIAS IN THE CONSTRUCT VALIDITY OF THE JAT

A test's construct validity refers to the extent to which the test measures a theoretical construct or feature such as intelligence, mechanical insight or aptitude for music (Anastasi 1976). This reasonably simple definition hides an extremely complex theme, as can be seen from the fact that Messick (1980) distinguishes about twelve different types of this validity.

The fact that a test's predictive validity for a minority group is satisfactory enough is no guarantee that the same can be said for its construct validity (Bernal 1975). The two types of validity are determined in completely different ways. In contrast to predictive validity, where the correlation between a test score and a criterion score is regarded as an adequate indication of validity, Cleary *et al.* (1975) regard the evaluation of construct validity as a complex combination of logical and empirical actions on the basis of which both the test and its underlying theory (in other words the construct) are evaluated at the same time. This suggests that it is essential when determining construct validity to evaluate several internal aspects of the test. It is also important to note that none of the internal criteria as such is adequate for making a categorical statement about the construct validity of the test. In this regard Jensen (1980) also points out that a test score is merely an attempt at measuring a specific construct. Nevertheless, together the internal characteristics (criteria) of a test provide an indication of the extent to which the test or test battery succeeds in measuring a specific construct.

6.1 DEFINITIONS OF BIAS IN THE CONSTRUCT VALIDITY OF A TEST

Reynolds (1983) defines bias in construct validity as follows:

> *Bias exists in regard to construct validity when a test is shown to measure different hypothetical traits (psychological constructs) for one group than for another or to measure the same trait but with differing degrees of accuracy* (p. 245).

This definition contains a general assumption that can be studied from various points of view. Several writers including Green (1972), Petersen (1980), Bond (1981), Cole (1981), Shepard (1981) and Sundberg and Gonzales (1981) mention that bias in construct validity means that a test measures one thing in a certain group and another thing in another group while the assumption is made that the same construct is being measured. Poortinga (Van der Flier & Drenth 1980) refers to the above

as functional equivalence, and according to him it is only one of three possible aspects of the comparability of the test scores of different groups. The other two are score equivalence (the construct or ability should be measured on a comparable scale for the groups; Hui & Trandis (1985, p.135) refer to it as "scalar equivalence") and item equivalence (the relations between the items should be similar for the various groups).

In the definition of construct validity by Jensen (1980) another element is added to those mentioned above. According to him bias occurs when there are systematic errors that are associated with the group membership of the individuals.

As far as the incidence of construct bias is concerned Scheuneman (1981a) maintains that although tests are basically valid (absence of bias in construct validity), bias is often found in the form of an underestimate of the abilities of minority groups. The latter assumption is supported to a certain extent by the conditional probability definition of Loyd (1983):

> *When individuals from different groups (cultural, racial, sexual, etc.) with the same level of ability have different probabilities of success on a particular test, the test is said to be biased* (p. 3).

Scheuneman and Loyd's assumptions therefore imply that although tests basically measure the same constructs in different groups, bias means that the total test score provides an underestimate (owing to systematic errors that have a bearing on the group membership of the individuals) of the"real" abilities of certain minority groups.

6.2 CONSTRUCT VALIDITY OF THE JAT FOR WHITE, INDIAN AND BLACK PUPILS

If the JAT measured the same constructs among the Indian and the black testees as it did among the white testees, in other words the battery were unbiased with regard to construct validity, then

- similar test reliabilities,
- similar rank order correlations of item difficulty values, and
- the same factors

would be found for the three population groups.

The results for each of the above matters will be discussed subsequently.

6.2.1 Test reliabilities

The reliability (internal consistency) of the tests of the JAT were calculated with the Kuder-Richardson Formula 20 (K-R20). These values were based on all the items that had been answered by the testees, in other words the item data that had been obtained with Time 3. The significance of the differences between the reliabilities was calculated with the equation

$$F = \frac{1 - r_{tt_1}}{1 - r_{tt_2}}$$

where r_{tt} is the reliability of the test. This equation is based on the method that was suggested by Feldt (Reynolds 1982, p.209). The quantity $(1-r_{tt})$ represents the error variance term. If a significant value is obtained, the test should be regarded as biased

> *...since errors due to domain sampling are different across groups creating disparities in the ability of the test to measure the trait in question for all groups* (Reynolds 1982, p.209).

The results appear in Table 6.1. They reveal that, as far as the JAT is concerned, similar test reliabilities were found among white and Indian testees. The only exception is the JAT 10 (however owing to the large number of degrees of freedom that are involved all the *F* values were statistically significant at the 1 % level). In the case of the black testees low test reliabilities were found for the JAT 2, JAT 4 and JAT 10 in particular and this leads to doubt as to whether these tests measure the same traits among white and black testees. As far as the SPM is concerned there was very little difference between the reliabilities of the three population groups.

From the above it is clear that, with the exception of one test among the Indian and three tests among the black testees, similar test reliabilities were found for the three population groups.

TABLE 6.1
SIGNIFICANCE OF THE DIFFERENCES (F) BETWEEN THE TEST RELIABILITIES (K-R20) OF THE POPULATION GROUPS WITH REGARD TO THE JAT AND SPM

Tests	White	Indian	Black	Differences between:	
	(N=1 056)	(N=1 063)	(N=1 093)	White/Indian	White/Black
	K-R20	K-R20	K-R20	F	F
JAT 1: Classification	0,632	0,648	0,637	1,05*	1,01
JAT 2: Reasoning	0,794	0,782	0,325	1,06	3,28
JAT 3: Number Ability	0,802	0,801	0,658	1,01	1,73
JAT 4: Synonyms	0,783	0,810	0,292	1,14	3,26
JAT 5: Comparison	¹	¹	¹	-	-
JAT 6: 2-D	0,872	0,873	0,826	1,01	1,36
JAT 7: 3-D	0,850	0,873	0,804	1,18	1,31
JAT 8: Memory (Paragraph)	0,802	0,727	0,658	1,38	1,73
JAT 9: Memory (Symbols)	0,865	0,839	0,835	1,19	1,22
JAT 10: Mechanical Insight					
Boys	0,763	0,627	0,326	1,57	2,84
Girls	0,681	0,514	0,293	1,52	2,22
Total	0,748	0,603	0,318	1,58	2,71
SPM	0,847	0,899	0,930	1,52	2,19

NOTE: ¹K-R formulas not applicable to tests where speed plays a major role
* All F values significant ($p < 0,01$)

6.2.2 Rank order of item difficulties

To determine the degree of similarity between the groups in the order of item difficulties the items were arranged in ascending order of difficulty and Spearman's rank order correlation (*rho*) was calculated between the sets of difficulty values (white/Indian amd white/black). By means of the value $1-rho^2$ an estimate was obtained of the proportion of the groups x items interaction variance which can be attributed to the disordinal aspect of the groups x items interaction. The results appear in Table 6.2.

TABLE 6.2

RANK ORDER CORRELATION (SPEARMAN RHO) BETWEEN THE ITEM DIFFICULTY VALUES FOR THE JAT FOR WHITE/INDIAN AND WHITE/BLACK PUPILS, AND PROPORTION OF GROUPS X ITEMS INTERACTION VARIANCE ($1-RHO^2$)

Tests	White/Indian		White/Black	
	rho	$1-rho^2$	rho	$1-rho^2$
JAT 1: Classification	0,80	0,36	0,66	0,56
JAT 2: Reasoning	0,94	0,12	0,80	0,36
JAT 3: Number Ability	0,98	0,04	0,95	0,10
JAT 4: Synonyms	0,51	0,74	0,45	0,80
JAT 6: 2-D	0,93	0,14	0,87	0,24
JAT 7: 3-D	0,93	0,14	0,80	0,36
JAT 10: Mechanical Insight	0,86	0,26	0,55	0,70

As the majority of the correlations in Table 6.2 are smaller than 0,90 (in other words they are not virtually perfect correlations), the results indicate an interaction between population groups and item difficulty values. In the case of the JAT 4 the values are particularly large and 74 % of the interaction variance among the Indian and 80 % among the black testees are attributed to the disordinal aspect of the groups x items interaction. The general conclusion is that for most of the tests the order of item difficulties are not similar for the three population groups.

6.2.3 Factorial similarity

To determine whether the tests of the JAT measure the same factors among the white, Indian and black testees the intercorrelations (Pearson product moment) of the ten test scores - that had been obtained with the standard testing time - were calculated separately for each population group. The intercorrelations appear in Table 6.3. In order to determine whether the results of the three population groups could be combined and regarded as one sample for the purpose of factor analysis the different correlation matrices were compared with the aid of the Box test (Lawley & Maxwell 1963), which converts the correlation matrices into covariance matrices. The results revealed that the covariance matrices of the Indian and black testees differed significantly ($p< 0,001$) from those of the white testees and that the data could therefore not be pooled. Consequently it was decided to perform the factor analysis for the various groups separately.

A principal component analysis (BMD programme) was then carried out on the intercorrelations of Table 6.3 to determine the number of factors that could be extracted for each individual group. Table 6.4 reveals that for each of the three groups there are two components of which the eigen values are greater than 1,0. A third component is also relatively large, although it does not reach the margin of $\geq$ 1,0. Kim and Mueller (1978) point out in this regard that the above margin is merely an indication of the number of factors that could be extracted and that it should definitely not be regarded as a fixed "law".

TABLE 6.3

INTERCORRELATIONS OF THE TEN TEST SCORES OF THE JAT FOR WHITE, INDIAN AND BLACK PUPILS

NOTE: (a) White pupils (N=1 056)

Tests	1	2	3	4	5	6	7	8	9	10
JAT 1: Classification	1,00									
JAT 2: Reasoning	0,35	1,00								
JAT 3: Number Ability	0,22	0,63	1,00							
JAT 4: Synonyms	0,32	0,61	0,42	1,00						
JAT 5: Comparison	0,13	0,36	0,45	0,27	1,00					
JAT 6: 2-D	0,39	0,47	0,34	0,34	0,25	1,00				
JAT 7: 3-D	0,45	0,49	0,32	0,37	0,23	0,71	1,00			
JAT 8: Memory (Paragraph)	0,24	0,50	0,42	0,45	0,38	0,23	0,23	1,00		
JAT 9: Memory (symbols)	0,19	0,42	0,37	0,34	0,31	0,25	0,27	0,52	1,00	
JAT 10: Mechanical Insight	0,35	0,52	0,35	0,37	0,21	0,53	0,51	0,31	0,22	1,00

NOTE: (b) Indian pupils (N= 1 063)

Tests	1	2	3	4	5	6	7	8	9	10
JAT 1: Classification	1,00									
JAT 2: Reasoning	0,42	1,00								
JAT 3: Number Ability	0,30	0,64	1,00							
JAT 4: Synonyms	0,33	0,67	0,51	1,00						
JAT 5: Comparison	0,20	0,40	0,40	0,41	1,00					
JAT 6: 2-D	0,49	0,44	0,32	0,27	0,21	1,00				
JAT 7: 3-D	0,51	0,48	0,34	0,31	0,22	0,73	1,00			
JAT 8: Memory (Paragraph)	0,16	0,47	0,38	0,49	0,31	0,16	0,18	1,00		
JAT 9: Memory (symbols)	0,22	0,44	0,40	0,42	0,31	0,27	0,30	0,43	1,00	
JAT 10: Mechanical Insight	0,33	0,42	0,34	0,38	0,23	0,45	0,44	0,25	0,29	1,00

TABLE 6.3 (continued)

Note: (c) Black pupils (N=1 093)

Tests	1	2	3	4	5	6	7	8	9	10
JAT 1: Classification	1,00									
JAT 2: Reasoning	0,29	1,00								
JAT 3: Number Ability	0,26	0,29	1,00							
JAT 4: Synonyms	0,23	0,33	0,27	1,00						
JAT 5: Comparison	0,28	0,27	0,43	0,25	1,00					
JAT 6: 2-D	0,37	0,28	0,26	0,21	0,30	1,00				
JAT 7: 3-D	0,38	0,30	0,21	0,18	0,22	0,64	1,00			
JAT 8: Memory (Paragraph)	0,24	0,28	0,24	0,24	0,32	0,18	0,16	1,00		
JAT 9: Memory (symbols)	0,22	0,21	0,28	0,15	0,28	0,25	0,23	0,35	1,00	
JAT 10: Mechanical Insight	0,26	0,24	0,23	0,22	0,25	0,23	0,21	0,23	0,19	1,00

TABLE 6.4

PRINCIPAL COMPONENT ANALYSIS OF THE TESTS OF THE JAT FOR WHITE, INDIAN AND BLACK PUPILS

Component	White		Indian		Black	
	Variance explain*	Cumulative proportion of total variance	Variance explain	Cumulative proportion of total variance	Variance explain	Cumulative proportion of total variance
1	4,39	0,44	4,39	0,44	3,41	0,34
2	1,38	0,58	1,45	0,58	1,18	0,46
3	0,81	0,66	0,74	0,66	0,92	0,55
4	0,72	0,73	0,67	0,73	0,82	0,63
5	0,66	0,79	0,66	0,79	0,80	0,71
6	0,54	0,85	0,58	0,85	0,68	0,78
7	0,52	0,90	0,53	0,90	0,65	0,85
8	0,41	0,94	0,44	0,95	0,64	0,91
9	0,30	0,97	0,27	0,97	0,54	0,96
10	0,27	1,00	0,27	1,00	0,35	1,00

Note: *The variance explained for each component is the eigen value for that component.

A further test that was carried out, namely the Scree test of Cattell (Kim & Mueller 1978, p. 44), revealed that no more than three factors could be extracted. In the light of the above a three-factor solution was decided on.

Factor analysis according to the principal factor method (BMD) with initial communalities that are equal to R^2 was carried out on the correlations in Table 6.3. Three factors were extracted for each population group and by means of the Varimax method rotated to simple structure. The factor loadings of the rotated factors appear in Table 6.5 (the highest loadings for each factor are underlined).

As can be seen from Table 6.5 Factor 1 has high loadings (only loadings of 0,30 and higher were taken into consideration) with respect to Spatial 2-D and 3-D for the white and Indian as well as the black pupils, while moderate loadings were found with respect to the white and Indian pupils for Mechanical Insight and the two reasoning tests, the JAT 1 and JAT 2, as well. As the highest loadings were with respect to the 2-D and 3-D, this factor could perhaps be identified as a <u>spatial</u> factor.

Factor 2 is mainly a <u>memory</u> factor, as can be seen from the loadings that were found with respect to all three population groups for the JAT 8 and JAT 9. However this factor also contains elements that differ with respect to the three population groups: the white pupils have a loading for the JAT 2 that was not found with respect to the other two groups, while loadings for the JAT 3 and JAT 5 were found with respect to the white and the black pupils but not with respect to the Indian pupils. Perceptual speed also played a a role in the case of some groups. Factor 2 is therefore more comprehensive than a mere memory factor.

Factor 3 displayed loadings with respect to all three population groups for the JAT 2 and JAT 4. As language is a common element in these two tests, this factor was identified as a <u>verbal</u> factor.

At first glance it therefore seems as if the same factors were identified with respect to the three population groups. To study the similarity of the factor patterns more precisely, congruency coefficients (Harman 1967) and an analysis programme by Browne, Schoeman and Glas (1979) were used. The congruency coefficients of the factors appear in Table 6.6.

TABLE 6.5

FACTOR LOADINGS AND COMMUNALITIES (h^2) OF THE TESTS OF THE JAT FOR WHITE, INDIAN AND BLACK PUPILS

Factors	White				Indian				Black			
	1	2	3	h^2	1	2	3	h^2	1	2	3	h^2
JAT 1: Classification	0,451	0,095	0,269	0,285	0,540	0,062	0,270	0,368	0,368	0,242	0,325	0,299
JAT 2: Reasoning	0,414	0,497	0,550	0,721	0,378	0,274	0,746	0,775	0,222	0,220	0,484	0,332
JAT 3: Number Ability	0,248	0,596	0,274	0,492	0,254	0,293	0,589	0,497	0,135	0,418	0,362	0,324
JAT 4: Synonyms	0,284	0,349	0,561	0,517	0,195	0,374	0,660	0,614	0,097	0,152	0,542	0,326
JAT 5: Comparison	0,147	0,618	0,010	0,403	0,139	0,298	0,398	0,267	0,155	0,478	0,333	0,363
JAT 6: 2-D	0,819	0,215	0,051	0,719	0,839	0,119	0,119	0,732	0,705	0,206	0,195	0,578
JAT 7: 3-D	0,808	0,174	0,141	0,702	0,832	0,146	0,154	0,738	0,832	0,115	0,147	0,728
JAT 8: Memory (Paragraph)	0,101	0,588	0,383	0,502	0,051	0,548	0,390	0,455	0,076	0,476	0,274	0,308
JAT 9: Memory (Symbols)	0,140	0,518	0,250	0,350	0,204	0,569	0,273	0,440	0,175	0,555	0,087	0,347
JAT 10: Mechanical Insight	0,557	0,204	0,298	0,440	0,452	0,213	0,283	0,330	0,173	0,242	0,326	0,195
	2,201	1,847	1,084		2,201	1,098	1,915		1,493	1,182	1,124	
% variance explained	22,01	18,47	10,84		22,01	10,98	19,15		14,93	11,82	11,24	

TABLE 6.6

COEFFICIENTS OF CONGRUENCE FOR A THREE-FACTOR SOLUTION FOR WHITE, INDIAN AND BLACK PUPILS

Population group	Factor 1	Factor 2	Factor 3
White/Indian	0,99	0,94	0,94
White/Black	0,95	0,96	0,92

As coefficients of congruence of 0,90 and higher generally signify factorial similarity (Powers & Jones, 1984), it can be concluded on the basis of Table 6.6 that the same factors are measured in the three population groups.

An analysis programme by Browne *et al.* (1979) provided additional support for the preceding finding. This programme contains Browne's (1972) method of rotation to a least squares fit given a partially specified target. The rotated factor matrices of the three population groups (Table 6.5) served as the initial values in the programme. A mean factor matrix, as well as three "best fit" matrices (one for each population group), was subsequently obtained for the three population groups together. The residual matrices were obtained by subtracting the mean factor matrix from the best fit factor matrices. The residual matrices for the white, Indian and black testees are reported in Table 6.7. A residual is arbitrarily put at $\leq$ 0,15 to serve as a criterion for the acceptance of the assumption that the same factors occurred in the different population groups. If one takes this residual size as the criterion, it seems that all the residuals are of an acceptable size. Additional support is therefore given to the conclusion drawn with regard to the coefficients of congruence. The fact that the mean square residuals vary only between 0,052 and 0,058 implies that the model fits the data reasonably well.

To conclude it can be stated that the JAT to a large extent measured the same constructs among the white, Indian and black pupils in the sample.

TABLE 6.7

RESIDUAL MATRICES FOR WHITE, INDIAN AND BLACK PUPILS

Factors	Factor 1			Factor 2			Factor 3		
Tests	White	Indian	Black	White	Indian	Black	White	Indian	Black
JAT 1: Classification	-0,059	0,043	-0,009	-0,050	-0,030	0,140	-0,028	-0,040	0,020
JAT 2: Reasoning	0,002	0,002	0,007	0,080	0,025	-0,064	-0,051	0,008	-0,081
JAT 3: Number Ability	0,015	0,011	-0,010	0,039	-0,085	0,001	-0,062	0,075	-0,017
JAT 4: Synonyms	0,007	-0,015	0,044	-0,000	0,127	-0,052	-0,054	-0,076	-0,001
JAT 5: Comparison	0,023	-0,032	0,054	0,011	-0,140	0,039	-0,100	0,086	0,113
JAT 6: 2-D	0,010	-0,004	-0,038	0,002	-0,024	0,016	0,008	0,018	0,041
JAT 7: 3-D	-0,052	-0,042	0,051	0,018	0,043	-0,062	0,047	0,016	-0,030
JAT 8: Memory (Paragraph)	-0,013	-0,019	0,038	-0,024	0,004	-0,048	0,100	-0,042	-0,017
JAT 9: Memory (Symbols)	-0,053	0,034	-0,015	-0,089	0,010	-0,038	0,144	0,023	-0,067
JAT 10: Mechanical Insight	0,106	0,029	-0,147	-0,042	0,015	0,070	0,008	-0,064	0,027

NOTE: The square root of the mean of the squared residuals is as follows:
White : 0,056
Indian : 0,052
Black : 0,058

7.0 ITEM BIAS IN THE JAT

7.1 DEFINITIONS OF ITEM BIAS

In the simplest definition of item bias, bias is compared to group differences as far as the mean of the item score is concerned (Durovic 1975). Thus, if the index of the item difficulty value is the same for the two population groups concerned, the item is not biased. According to Hambleton, Swaminathan, Cook, Eignor and Gifford (1978) the problem with this definition or method is that it describes not only the item, but also the group that was tested. In their opinion a usable index for detecting bias should be independent of the group to whom the test was administered. An additional problem with this definition concerns the fact that there is little support for assuming the equality of groups (Burrill & Wilson 1980; Kelley 1982; Reynolds 1982). Even if it is assumed that the different groups all have the same abilities or potential, it still cannot be taken for granted that the groups all had equal opportunities to obtain the required knowledge and skills. Jensen (1980) refers to this problem as the so-called egalitarian fallacy. The egalitarian philosophy states that all human populations are identical with regard to all cognitive properties or abilities. In support of Jensen, Reynolds (1982) points out that there is no *a priori* basis for the claim that one group does not differ from another as far as cognitive abilities are concerned. Shepard (1982) summarizes the viewpoint of those who oppose the egalitarian philosophy as follows:

> *An a priori assumption of equivalent group means, however, has been rejected by most scholars who believe that the existence of a difference between groups is not automatically a sign of bias* (p. 13).

The preceding definitions of item bias are based on characteristics of the classical test theory and are unconditional as far as assumptions concerning the ability level of testees are concerned. However more recent definitions of item bias are increasingly based on the item response theory (latent trait theory) (Humphreys 1986). An important characteristic of these definitions is that a condition is imposed regarding the ability level of the testees: only testees with the same ability level are compared with one another (Mellenbergh 1983). A typical definition of item bias based on the item response theory (hereafter referred to as IRT) reads as follows:

> ...*an item is generally considered biased if equally able members* (my underlining) *of different groups have unequal chances of success on the item* (Subkoviak, Mack, Ironson & Craig 1984, p. 49).

This view of item bias is presently supported by most researchers in this field, for example Petersen (1980), Scheuneman (1980, 1981a), Marascuilo and Slaughter (1981), Shepard, Camilli and Averill (1981), Rudner and Getson (1982), Mellenbergh (1983), and Rulick and Dorans (1984).

Definitions and descriptions of item bias that are based on IRT methods for their identification can be summarized as follows:

- An item is biased against members of a group if the expected performance on the item is lower for persons in that group than for persons of a similar level of ability in the other group.

- Bias in this context may be the result of multidimensionality, in other words, the probability of answering an item correctly depends on more than one latent trait and therefore the item does not measure the same characteristic in the different groups (Linn, Levine, Hastings & Wardrop 1981; Shepard 1981).

- Item bias is a contextual property, an anomaly within the context of other items, which leads to the fact that bias cannot be identified in an isolated item - test items designed to measure the same construct should consequently be studied together (Shepard 1982).

- In investigations into item bias evidence is sought for an interaction between performance on an item and group membership while differences in the levels of ability of the groups are held constant (Traub & Lam 1985).

7.2 THE DETECTION OF ITEM BIAS

The statistical or psychometric problem in the detection of item bias involves finding methods that will indicate those differences in the test performance of groups that represent a real difference (a difference that has to be recognized as such) and those differences that are merely the result of bias (for example owing to the way in which the item was formulated)(Burrill & Wilson 1980). Techniques for detecting item bias therefore attempt to identify specific test items that are associated with cultural differences and that contribute to inaccurate measurement (Rudner & Getson 1982).

A shortcoming encountered in all these methods is that they are based on an internal test criterion. This inevitably introduces an element of circularity to the logic of the methods for detecting bias. The total score of the test (to which items that may perhaps be biased actually contributed) is used to identify persons with equal abilities and the performance of these persons is subsequently used to identify biased items (Shepard, Camilli & Williams 1983; Traub & Lam 1985). Thus the methods can only succeed in detecting relative bias and not absolute bias.

The following are some of the most important methods or techniques for the detection of biased items:

- Discrimination value of the item
- Rank order of item difficulty values
- Transformed item difficulty values
- Analysis of variance
- Analysis of the responses to item distractors
- The chi-square
- Methods based on item response theory.

Since the latter two methods were used in this investigation, they are discussed in greater detail in the following paragraphs.

7.2.1 The chi-square

The chi-square approach (χ^2) to the identification of biased items examines the probability that testees from different groups who have the same level of ability will answer the same item correctly (Osterlind 1983). If an item is unbiased, individuals of the same ability level should also have the same probability (chance) of answering the item correctly, regardless of the group to which the individual belongs (Scheuneman 1979). This approach therefore avoids the dependence of items X group interaction (as in the case of analysis of variance) for the identification of biased items. Two methods that depend on the chi-square approach and are often used in investigations (in this investigation as well) are Scheuneman's chi- square and the "full" chi-square.

(a) Scheuneman's chi-square technique

According to Scheuneman (1979, 1980) the chi-square technique that she proposes should be regarded as a variation of the item response theory (latent trait procedures). Both the difficulty and discrimination values of the items are allowed to vary, but the continuous curve which is characteristic of the latent trait procedure is abandoned. In its place the test total score is divided into a small number of discreet ability categories (three to five) and the performance within these categories is compared for two or more cultural groups. Thus Scheuneman's chi-square technique involves the use of the total raw score of a homogeneous test to measure or indicate ability. The distribution of the total score is divided into three to five categories or intervals that are each defined by a certain range of the total score (i.e. if the test consists of 30 items counting one point each, the five ability categories may include testees who obtained total scores ranging for example from 0-9, 10-15, 16-20, 20-25, 26-30). The distribution of the number of correct responses within each ability category is obtained for each item and for every population group. The number of correct responses within each ability category represents the obtained frequency used in the chi-square test. The proportion of correct responses of all the testees (regardless of their ethnic background) whose scores fell within a particular ability category is used to determine the expected frequency. In other words, the expected value for each cell (E_{ij}) is obtained by multiplying the proportion of all testees who answered the item correctly and whose total score fell inside interval j by the number of testees within the cell. Thus

$$E_{ij} = \frac{O_{\cdot j}}{N_{\cdot j}} N_{ij}$$

where

$O_{\cdot j}$ = the number of testees in the total score interval (ability category) j who answered the item correctly
$N_{\cdot j}$ = the total number of testees in the score interval j, and
N_{ij} = the total number of testees in group i and in score interval j.

The test statistic (χ^2) is

$$\chi^2 = \Sigma \frac{(B_e - B_o)^2}{B_e} + \Sigma \frac{(W_e - W_o)^2}{W_e}$$

where

B_e = the expected frequencies of Group B
B_o = the obtained frequencies of Group B
W_e = the expected frequencies of Group W, and
W_o = the obtained frequencies of Group W.

One of the shortcomings of this method (as well as of the "full" chi-square) concerns the fact that when there are a considerable number of biased items in a test, the test total score will also be biased. Consequently the ability categories, which have been formed on the basis of the total score, will not reflect the true potential of the group concerned. When there is a considerable difference between the total score distribution of the two groups concerned, for example as a result of the fact that the one group finds the test much easier than the other group, the establishment of comparable ability groups poses a practical problem.

Intasuwan (1979) and Baker (1981), among others, raised a more serious objection to Scheuneman's technique in particular. In their opinion the fact that she does not incorporate the proportion of incorrect answers in her technique results in the test statistic which she uses not really having a chi-square distribution. In other words it is only "half" a chi-square.

(b) The full chi-square

Camilli (Shepard, Camilli & Averill 1981) indicated that the results of a chi-square statistic that are based on the proportion of correct responses (p) differ from those in which the proportion of incorrect responses (q) are used. This was an awkward finding, since the proportions of correct and wrong answers are supposed to reflect the same information. Camilli, as well as Intasuwan (1979), consequently suggested that the full chi-square, which includes both the correct and the wrong responses of the testees, be computed. In other words, a conventional chi-square for a two-by-two contingency table (for example ethnic group X correct-incorrect responses) should be computed for each ability level and summed across all the ability levels. Thus the calculation procedures match those of the Scheuneman technique, with the exception that the incorrect responses are also taken into account. If the score for the item involved does not form part of the total score, the test statistic calculated for this item can be regarded as a true chi-square (Intasuwan 1979). However this implies that score intervals (ability groups) have to be determined for each item in the test. According to Rudner and Convey (1978) deviating from this approach, in other words keeping the ability groups the same for all the items in a specific test and including the item under scrutiny in the total score, has the following particular advantage:

> *The inflation of the χ^2 values will be systematic when identical intervals are used for each item ... This systematic inflation allows the χ^2 to be used as a relative index of bias* (p. 20).

Although she is aware of criticism , Scheuneman (1981b) still prefers her amended chi-square method because it is far more difficult to create cells of adequate size with the full chi-square in the case of items that are particularly easy. It can also be argued that the need for a significance test (and consequently for a distribution with known char-

acteristics) is subordinate to the need for an index of bias. Shepard (1981) also points out that the final conclusions about bias should be based on logical grounds and that the deviation of the item - compared with the other items measuring the same construct - should weigh more than the power of the statistical test.

As far as the present investigation is concerned, the chi-square tests (both Scheuneman's and the full chi-square) were used to identify biased items in the JAT (Tests 1, 2, 3, 4, 6, 7 and 10). For all the tests except the two spatial tests (JAT 6 and JAT 7), the ability groups were formed on the basis of the total score on the Raven SPM. After the score distributions of the three population groups in the latter test had been studied, four ability groups were decided on. Table 7.1 shows the ability groups according to range of total scores on the SPM, as well as the number of testees in each group. This classification of ability groups applies to all the items except those of the JAT 6 and JAT 7. For these tests the ability groups were formed on the basis of the test total score involved. (This exception is based on the assumption that the total scores of the spatial tests are not only unbiased, but that they also give a better indication of the specific ability than does the SPM.)

TABLE 7.1

ABILITY GROUPS ACCORDING TO RANGE OF TOTAL SCORES ON THE SPM, AND NUMBER OF TESTEES IN EACH ABILITY GROUP

Ability groups: Total score on SPM	White	Indian	Black
	N	N	N
44 - 60	716	514	63
39 - 43	208	252	139
34 - 38	80	159	166
00 - 33	52	138	725
TOTAL	1 056	1 063	1 093

Since comparisons are drawn between two population groups and four ability categories every time, the total degrees of freedom for the two chi-square techniques are as follows:

Scheuneman technique

df = (K - 1) (r - 1), where K is the number of population groups and r the number of ability groups
= (2 - 1) (4 - 1)
= 3 (indicated as χ^2_3 in Appendix B)

Full chi-square (the number of both correct and incorrect answers are taken into account):

df = s(g - 1), where g is the number of population groups and s the number of ability groups
= 4(2 - 1)
= 4 (indicated as χ^2_4 in Appendix B)

The results of the bias analysis appear in Appendix B.

7.2.2 The item response theory

The mathematics involved in item response theory (IRT) is rather complex and this theory will therefore not be discussed in detail. Comprehensive descriptions can be found *inter alia* in Hambleton, Swaminathan, Cook, Eignor and Gifford (1978), Warm (1978), Hambleton (1980), Lord (1980), and Hulin, Drasgow and Parsons (1983).

The item response theory is based on the idea that latent traits (hypothetical dimensions of cognitive abilities) offer an explanation of the coherence among different items of a test (Hambleton & Cook 1977; Van der Flier 1980). Because the traits are not directly measurable and are therefore "imperceptible", they are often referred to as latent traits. According to Carter and Wilkison (1984) latent traits can be regarded as psychological dimensions underlying the abilities and attitudes of individuals. Performance in a task related to a trait can be predicted by estimating an individual's ability level with regard to the trait and using these estimated scores to predict performance.

The complete latent space is the space that is determined by all dimensions required to explain the coherence among the items (Van der Flier 1980). According to its definition, the conditional distribution of the item scores within this space (given the position of the latent trait) is equal for all relevant populations. When a test consists of homogeneous items, it is assumed that the coherence among the items can be explained by a single underlying trait. The regression of the observed item score on this trait (θ) is called the item characteristic curve (ICC) or item ogive (Traub & Lam 1985).

The advantages of the IRT over the classical test model can be summarized as follows:

- The testee's ability is estimated on the same ability scale as any subgroup of items fitting the model.
- The item parameters are invariant across subgroups of testees and consequently render it possible to draw comparisons between population groups.
- In the case of the IRT the scales actually have the properties of an interval scale (raw scores usually lead to an ordinal scale, but in order to be able to use certain statistical procedures, it was decided pragmatically in the classical test theory that the scale had the properties of an interval scale).

A disadvantage of the IRT involves the fact that no set of data will ever fully satisfy the assumptions made regarding any latent trait model. Apparently the latent trait models are also not robust with regard to deviations concerning the assumption of unidimensionality (Hambleton 1980).

A number of latent trait models have been developed. Most models presuppose a more or less s-shape item characteritic curve. The best-known models are the three-parameter normal ogive

model, the three-parameter logistic model, the two-parameter logistic model and the one-parameter logistic model (also known as the Rasch model). The first two models yield similar ICCs, but the logistic distribution function not only proves to be mathematically less complicated, but it is also used more often (Van der Flier 1980). In all these models it is assumed that an item is unbiased if ... *examinees of the same ability level, but of different group affiliations, have equal probabilities of responding correctly* (Rudner, Getson & Knight 1980, p. 222).

Since the three-parameter logistic model is also used in the present investigation, a discussion of the parameters of this model follows next.

The mathematical form of the item characteristic curve of the three-parameter logistic model is as follows (Hambleton *et al.* 1978, p. 473):

$$P_g(\theta) = c_g + (1 - c_g)\frac{e^{Da_g(\theta - b_g)}}{1+e^{Da_g(\theta - b_g)}}, \quad (g = 1,2,\ldots,n).$$

where

$P_g(\theta)$ = probability that a testee with ability θ will answer item g correctly,
D = a constant scaling factor (usually taken as 1,7),
e = the mathematical constant 2,71828...,
a_g, b_g and c_g = the parameters describing item g

The a parameter is the discriminating power of the item. The steeper the slope of the curve the better the item discriminates. Discrimination values of 0,5 to approximately 2,5 are reasonably typical while a value smaller than 0,5 is not adequate for testing purposes (Ree 1979).

The b parameter is an index of the degree of difficulty of the item and is represented by a point on the ability scale (θ) that coincides with the inflection point of the curve (i.e. the point where the curve changes direction). The further to the right the inflection point, the more difficult the item, and the more to the left, the easier the item. Although b usually lies between -3 and +3, the choice of this scale is arbitrary (Ree 1979).

The c parameter (pseudochance parameter) is the lower asymptote of the curve and specifies the probability of a testee of extremely limited ability answering the item correctly by means of guessing. This parameter is usually larger than 0,0 and smaller than 0,30.

Different methods can be employed to estimate the above item parameters (Ree 1979; Osterlind 1983). A method that is commonly used is the one contained in the computer program LOGIST, compiled by Wood, Wingersky and Lord (1976). This program not only estimates the item parameters (a, b and c) as well as the ability parameter (θ), but also places them on the same scale. When the parameters of the different population groups are estimated separately, they will not be identical since different samples are used in the parameterization.

Although the parameters of both groups lie virtually on the same scale, there is a linear transformation difference between them (Osterlind 1983). This difference is caused by the fact that θ is arbitrarily de-

fined by a mean of 0 and a standard deviation of 1 in each separate parametrication. In this regard Dorans (1979) declares that

> *... the fact that standard scores from separate derivation samples have identical means and variances does not mean that these scores are expressed in a common metric. Scores are expressed in a common metric only when they share a common reference point and a common unit* (p. 11).

Despite the fact that the a and b parameters are invariant from group to group, they are not invariant when the origin of θ changes arbitrarily for each parameterization. Consequently the scales of two separate groups have to be equated before the parameters can be estimated and compared (Dorans 1979). Various methods may be used for equating the θ scales (see for example Osterlind 1983; Stocking & Lord 1983; Lord & Wingersky 1984; Divgi 1985).

All that remains after the scales have been equated and the parameters estimated is to compare the item characteristic curves (ICCs) of the two groups with a view to identifying biased items. If the ICCs of the two groups are identical, the item concerned is unbiased. However if the ICCs differ, certain methods may be used to determine whether the deviation is such that the item should be considered biased. Rudner, Getson and Knight (1980) for example established an index of item bias based on the area between the ICCs of the two groups, while Lord (1980) developed a chi-square statistic for testing the equality of the a and b parameters of the groups (the c parameter is not tested because it was held constant while the a and b parameters were estimated).

The three-parameter logistic model is also used in the case of the JAT 1 and JAT 2. Instead of the scores of the particular tests, the test scores on the SPM were used to obtain estimates of the testees' ability traits (θs). The computer program LOGIST usually estimates the item parameters (a, b and c) and the ability parameter (θ) simultaneously by means of a series of iterative steps. However this also allows for the possibility of basing the ability estimates on another test than the one involved. In this case the θ scores of the SPM are entered, standardized ($\bar{x} = 0$, $S = 1$) and held constant in all subsequent phases of the estimation process. Since the θ s for the total group were estimated on the basis of the SPM, the θ s of the distinctive population groups were placed on the same scale. The c parameters (chance parameters) of the items of a particular test were estimated on the basis of the total group of testees and held constant. The a and b parameters of the items (in other words the discrimination and difficulty values) of the JAT 1 and JAT 2 are estimated separately for each population group and compared by means of the item characteristic curves (ICCs). The extent of bias in an item as reflected by the difference between the curves of two population groups (white pupils being the reference group) was determined by means of a method proposed by Linn, Levine, Hastings and Wardrop (1980). According to this method the area between two ICCs can be determined by dividing the area between $\theta = -3$ and $\theta = +3$ into 600 intervals and calculating the following four indices:

1. Base Group Higher Area: the size of the area between $\theta = -3$ and $\theta = +3$ in which the ICC of the base group (i.e. the reference group) lies above the ICC of the comparison group.

2. Base Group Lower Area: the size of the area between $\theta = -3$ and $\theta = +3$ in which the ICC of the base group lies below the ICC of the comparison group.

3. Absolute difference: the total of 1 and 2 above.

4. Square root of the Sum of Squares: the square root of the sum of the squared differences between the ICCs in the area $\theta = -3$ and $\theta = +3$.

An item that displays a large base group higher area (Index 1) but a small base group lower area (Index 2) is considered to be biased against the comparison group. Such a result implies that persons in the comparison group have a smaller probability of answering the item correctly than persons in the base group with the same estimated ability (θ). According to the authors mentioned earlier, items with a value of 0,2 and higher for Index 4 are probably biased. The values obtained by means of Index 4 are contained in Appendix B while the item characteristic curves of the JAT 1 and JAT 2 are given in Appendix C.

7.3 ITEM BIAS IN THE TESTS OF THE JAT

This section contains the findings regarding the items in the JAT 1, JAT 2, JAT 3, JAT 4, JAT 6, JAT 7 and JAT 10. As mentioned earlier, item bias was not examined for the JAT 5 (Comparison) and the two memory tests (JAT 8 and JAT 9), because performance in these tests is to a large extent a function of testing time. Both the JAT 1 (Classification) and JAT 2 (Reasoning) are extremely important as typical representatives of tests of general cognitive ability. For this reason indices of bias are not only determined by using Scheuneman's chi-square (χ^2_3) and the full chi-square (χ^2_4), but also item response theory (latent trait theory).

7.3.1 JAT 1: Classification

Before item bias as such was examined, the item *p* values (difficulty values) of the groups were compared by means of scattergrams. These particulars are reflected in Figure 7.1 and from them it seems that black testees found several items considerably more difficult than did their white counterparts. Testees from the two groups found reasonably few items to be approximately equally difficult (in other words items lying near or on the 45° line). In the case of the Indian pupils the situation was somewhat different. Despite the fact that they found the items generally more difficult than did the white testees, the *p* values were distributed fairly uniformly and few items lay particularly far from the 45° line.

According to Table A1 in Appendix A the mean percentage of testees who answered an item correctly was as follows: white - 78 %, Indian - 68 % and black - 56 %. The extent to which the difficulty values of the items differed for the three groups was also evident from the fact that in the case of the white testees the *p* values of 26 of the 30 items lay between 0,61 and 0,99, while for the Indian testees there were 20 and for the black testees only 14 such items.

The extent to which the poorer performance of the Indian and black pupils in the sample can be ascribed to bias is indicated in Table B1 (Appendix B). With the Scheuneman technique (χ^2_3) 13 items were identified

as ones that were probably biased against Indian pupils and 15 as ones that were probably biased against black pupils, while the index based on the item characteristic curve (ICC) identified 18 items that were biased against both the Indian and the black testees. Twelve of these items, namely Items 1, 2, 5, 8, 10, 11, 19, 22, 24, 26, 27 and 29 were common to both groups. The correspondence between the two methods of bias was reasonably large and with them 23 items (77 %) were classified as either biased or unbiased in the case of Indian testees and 25 items (83 %) in the case of black testees. The full chi-square (χ^2_4) yielded similar results and for this reason reference will be made mainly to the previous two indices. The item characteristic curves (ICCs) of the items for the white, Indian and black testees are given in Appendix C1. The fact that for some items the ICCs did not show the typical ogive form but that of a straight line points to a poor fit of the item response model with regard to the data. This may possibly be attributed to the multidimensionality of the item.

The following items for example displayed large indices of bias:
Indian testees: Items 8, 11, 13, 15, 22, 24, 26, 27 and 29
Black testees: Items 8, 11, 19, 20, 22, 24, 26, 27 and 29.

A common element in most of these items was that they presupposed knowledge or a degree of knowledge on the part of the testee, for example knowledge of tools (pliers, screws and nails; garden tools), instruments (electrical appliances; old-fashioned and modern lighting; magnifying instruments, such as a microscope; equipment for catching fish, mice and insects), and general knowledge (for example of different kinds of antelope and Western type of ladies' accessories).

Thus in the case of both the Indian and the black testees it seems that the single largest cause of bias lay in the fact that the pupils were not familiar with the objects represented by the pictures. Cultural and SES factors probably also played a role in this regard.

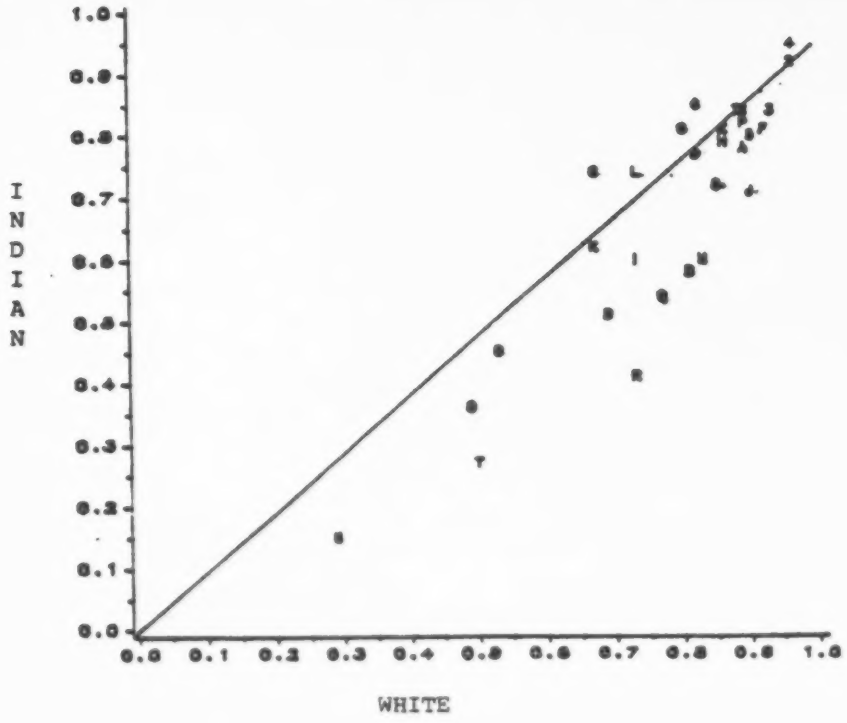
The above that may possibly be regarded as examples of true bias should be distinguished from apparent bias which stems from the way in which the testees sometimes answer an item. From the responses to Items 6, 7, 9, 10, 18 and 22 it seems that a larger percentage of the black than of the white or Indian pupils

. did not take all the details in an item into account, and
. chose the most illogical distractor.

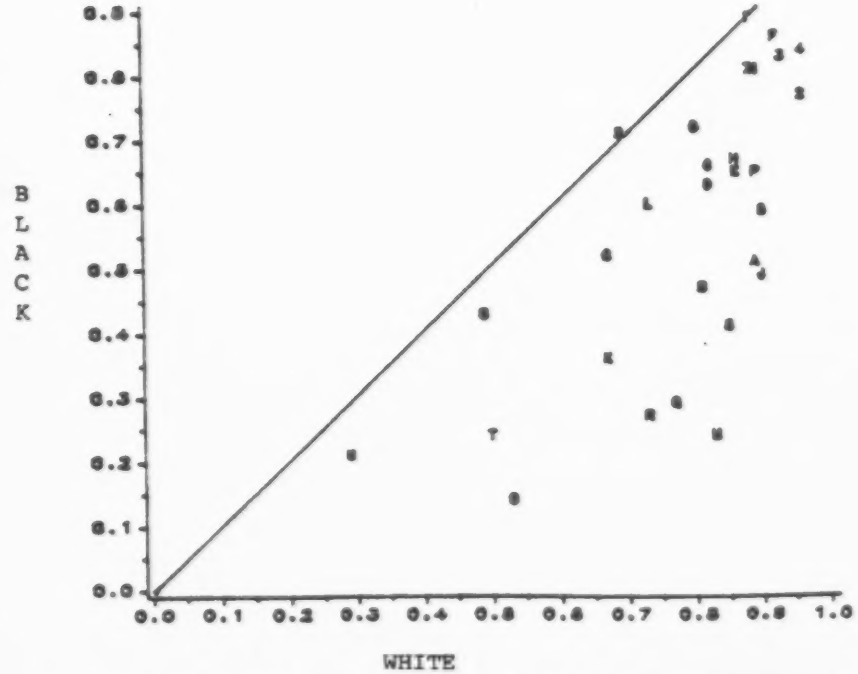
In the case of Item 22, for example, 26 % of the black pupils in the sample, compared with 6 % of the Indian and only 2 % of the white pupils, chose Distractor E which is one of the poorest of the five answer possibilities. In general it seemed that the item distractors were not equally appealing to white and black testees (see Table A1 in Appendix A).

To summarize: Although bias undoubtedly played a role in Indian and black pupils' performance in the JAT 1, it was not the only reason for the poorer performance of the black pupils in particular.

A) COMPARISON BETWEEN WHITE AND INDIAN PUPILS



B) COMPARISON BETWEEN WHITE AND BLACK PUPILS



CODE		ITEM
1	=	1
2	=	2
3	=	3
4	=	4
5	=	5
6	=	6
7	=	7
8	=	8
9	=	9
A	=	10
B	=	11
C	=	12
D	=	13
E	=	14
F	=	15
G	=	16
H	=	17
I	=	18
J	=	19
K	=	20
L	=	21
M	=	22
N	=	23
O	=	24
P	=	25
Q	=	26
R	=	27
S	=	28
T	=	29
U	=	30

FIGURE 7.1: JAT 1: SCATTERGRAM OF ITEM DIFFICULTY VALUES FOR WHITE/INDIAN AND WHITE/BLACK

7.3.2 JAT 2: Reasoning

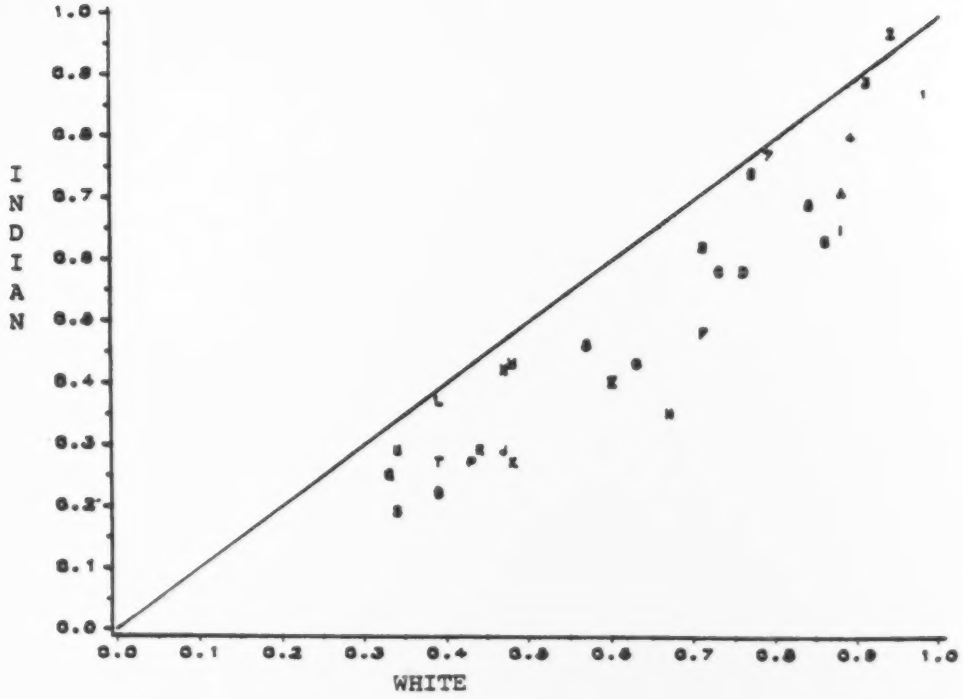
The scattergrams of the item difficulty values in Figure 7.2 reveal considerable differences between black and white testees in particular. In the case of the black testees only four items (1, 2, 4 and 9) had a *p* value of 0,5 and higher, while the *p* values of the vast majority of items were below 0,3. This means that these pupils found most items extremely difficult. The performance of the Indian pupils in the sample (in contrast to that of the black pupils) displayed considerably larger variation and in their case the items were distributed reasonably uniformly among low, average and high *p* values. Most of the items lay below the 45° line and were obviously more difficult for the Indian than for the white testees.

As far as bias is concerned, the results given in Table B2 in Appendix B indicate that with the χ^2_3 index only two items (Items 2 and 9) and with the ICC index only three (Items 2, 7 and 30) are unbiased for the black testees. It does not appear to be at all incidental that the items mentioned are more concerned with numbers than with words. In the case of the Indian pupils the χ^2_3 and ICC indices respectively classified 17 and 14 items as being possibly biased. Testees with the same ability - as determined with the Raven SPM - consequently do not have an equal chance of answering a verbal reasoning item correctly. Therefore, despite the fact that the SPM and JAT 2 do not measure exactly the same underlying dimensions of ability and a degree of scope should in fact be allowed for the interpretation of results, it is clear that the performance of the black pupils in the sample was affected more negatively by, inter alia, bias than the performance of their Indian counterparts.

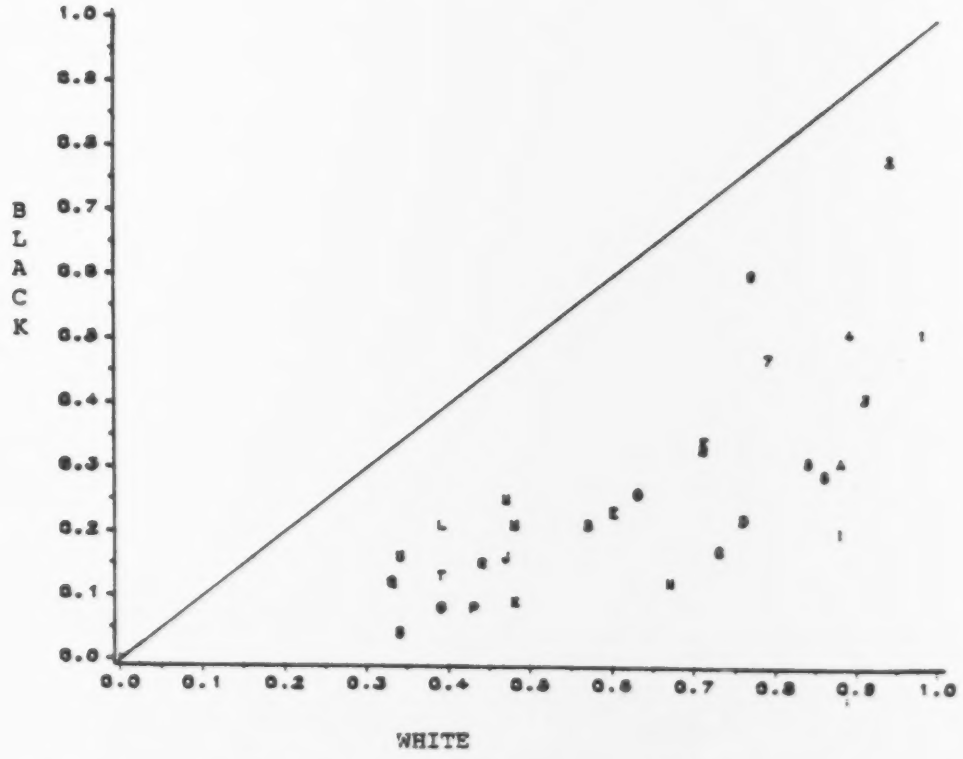
The ICCs of the items regarding the white, Indian and black testees are contained in Appendix C2. These figures show that in the case of four items (13, 18, 20 and 21) it was impossible to determine the ICCs of the black testees. The a, b and c parameters of these items deviated so radically that they could not be estimated at all.

In order to determine where the biggest problems occurred, the items of the JAT 2 were classified according to type and both the extent of the χ^2_3 values (Table B2 in Appendix B) and the difficulty values (Table A2 in Appendix A) were reviewed. The results appear in Table 7.2 and indicate that the black pupils experienced problems with all the item types. The most difficult type of item (lowest *p* values) was Shuffled Letters, while Number Series was the "easiest". The largest index of bias was found in the case of Verbal Analogies and Word Classification. Shuffled Letters was also the most difficult item type for the Indian pupils, while Word Classification, Number Series and Verbal Analogies were apparently easier. For these pupils Word Classification and Shuffled Letters yielded the largest bias indices.

A) COMPARISON BETWEEN WHITE AND INDIAN PUPILS



B) COMPARISON BETWEEN WHITE AND BLACK PUPILS



CODE		ITEM
1	=	1
2	=	2
3	=	3
4	=	4
5	=	5
6	=	6
7	=	7
8	=	8
9	=	9
A	=	10
B	=	11
C	=	12
D	=	13
E	=	14
F	=	15
G	=	16
H	=	17
I	=	18
J	=	19
K	=	20
L	=	21
M	=	22
N	=	23
O	=	24
P	=	25
Q	=	26
R	=	27
S	=	28
T	=	29
U	=	30

FIGURE 7.2 JAT 2: SCATTERGRAM OF ITEM DIFFICULTY VALUES FOR WHITE/INDIAN AND WHITE/BLACK

In an attempt to examine the cause of the problems further, the testees' responses to the item distractors were studied (Table A2 in Appendix A). Apart from the fact that the different groups did not find the distractors equally appealing, the following conclusions concerning the nature or cause of the problems can be reached on the basis of the response patterns for the different item types:

. Word Classification. Examples of this type of item are "Which word fits in least with the other four?" and "What goes best with a (object) and a (object)?"

Apart from a lack of specific word knowledge which possibly played a role in Items 15 and 20, as well as certain emotionally laden words such as "race" in Item 17 that attracted many black and Indian testees, it seems that the biggest problem was the fact that many pupils did not classify the objects on the basis of conceptual similarity, but on the basis of associative resemblances. In the case of for example Item 13, 78 % of the black testees, compared with 42 % of the Indian and 24 % of the white testees, chose an answer that suggested an associative resemblance; in the case of Item 6 the corresponding percentages were 61 %, 27 % and 14 % respectively. Poor performance can also be ascribed to lack of a critical or logical attitude. Apparently it did not occur to many testees when they chose a specific answer that other distractors might very well be correct answers too. For example, in Item 4, 33 % of the black, 14 % of the Indian and 8 % of the white testees chose Distractors C and D; yet if Distractor C (mother) were correct, Distractor D (aunt) would have been an equally possible answer.

. Number Series. Those series in which the increase is constant, for example 2; 4; 6; ?, cause remarkably few problems. However, as will be indicated later, it appears that these items are answered correctly for the wrong reasons. When the increase is not constant, for example 1; 3; 6; 10 (15) or in cases where the number series consists of a double series, in other words the 1st, 3rd and 5th numbers form one series and the 2nd, 4th and 6th numbers another series, for example 1; 8; 2; 7; 3; (6), many testees tend to ignore the preceding numbers and base their answers on the last two numbers only. This strategy is illustrated very clearly by, among others, Item 7. This item (in a somewhat amended form) reads as follows: 3; 6; 12; ? (24); 48. From the responses it is evident that 20 % of the black, 9 % of the Indian and 9 % of the white pupils simply subtracted 12 from 48. A further 17 % of the black, 7 % of the Indian and 6 % of the white pupils added up the two numbers preceding the question mark (6 + 12 = 18). In the case of Item 19, 40 % of the black, 35 % of the Indian and 24 % of the white testees considered only the difference between the last two numbers in the series; only 16 % of the black, compared with 29 % of the Indian and 47 % of the white testees noticed that the particular row of numbers consisted of a double series.

. Verbal Analogies. As in the case of the Number Series it seems that a considerably higher percentage of the black testees (compared with the Indian and white pupils) did not consider the analogy as a whole, but based their answers on the second part of it only. The more difficult the item, for example Item 26, the stronger was this tendency. In the case of Item 3 the fact that only 41 % of the black testees (compared with 89 % of the Indian and 91 % of the white pupils) answered the item correctly, can be attributed to the fact that as much as 35 % of the

TABLE 7.2
REASONING: RANGE OF CHI-SQUARE (χ^2_3) AND ITEM DIFFICULTY VALUES (p) ACCORDING TO ITEM TYPE

Item type	Number of item in test	Total number of items	Indian		Black	
			Range of χ^2_3 values	Range and mean p values	Range of χ^2_3 values	Range and mean p values
I Word Classification	1; 4; 6; 8; 13; 15; 17; 20	8	4,8-89,4	0,27-0,87 $\bar{x}$=0,56	35,4-187,7	0,09-0,51 $\bar{x}$=0,29
II Number Series	2; 5; 7; 16; 19; 23; 30	7	1,3-29,3	0,29-0,97 $\bar{x}$=0,55	1,7-41,4	0,16-0,78 $\bar{x}$=0,34
III Verbal Analogies	3; 18; 21; 26;	4	1,4-37,1	0,25-0,89 $\bar{x}$=0,54	30,2-253,5	0,12-0,41 $\bar{x}$=0,24
IV Arithmetical "story" problems	9; 11; 12; 26; 28; 29	6	0,7-34,1	0,19-0,74 $\bar{x}$=0,45	7,0-125,4	0,04-0,60 $\bar{x}$=0,23
V Linear Syllogisms	10	1	13,6	0,71	89,0	0,31
VI Shuffled Letters	24; 27	2	36,3-38,7	0,22-0,29 $\bar{x}$=0,26	53,8-70,1	0,08-0,15 $\bar{x}$=0,12
VII Knowledge	14; 22	2	1,5-27,5	0,40-0,43 $\bar{x}$=0,42	23,1-27,8	0,21-0,23 $\bar{x}$=0,22

black pupils (compared with 5 % of the Indian and 2 % of the white pupils) simply looked for a word associated with "grapes", namely "wine" (Distractor E). The differences between the groups are illustrated even more clearly by Item 18, where 88 % of the white, 65 % of the Indian and 20 % of the black testees answered the item correctly. Despite the fact that the item appeared to be reasonably easy - in amended form it read "Trolley is to Push as Canoe is to ....? (Row)" - it became a difficult item for the black testees because 38 % of them (compared with 21 % of the Indian and 3 % of the white pupils) ignored the first part and selected a word associated with Canoe (i.e. Float). Furthermore it was striking to note that, compared with the Indian and white testees, a larger percentage of the black pupils were attracted to the most improbable of the five answer options: in the case of for example Item 18, 12 % of the black, as against 1 % of, the Indian and onl 0,6 % of the white testees chose "fuel" as their answer. With regard to Item 26 a similar percentage of testees chose "direction" instead of "ruler" as a means with which to determine length.

. Arithmetical "story" problems. The following are some of the most common errors made by testees:

In answer to the question such as: "What is the smallest number that can be subtracted from ... in order to make it exactly divisible by 8?" (Item 9), 17 % of the black pupils, as against 6 % of the Indian and 5 % of the white testees did not indicate the number to be subtracted, but the number that would be divisible by 8. In a similar question, though one in which the testees had to indicate the smallest number to be added to a given number in order to make it exactly divisible by 6 (Item 11), 42 % of the black pupils - as opposed to 25 % of the Indian and 18 % of the white pupils - chose numbers that, if added to the given number, were not exactly divisible by 6. Another 16 % of the black, 6 % of the Indian and 4 % of the white testees simply chose the number divisible by 6, instead of the number to be added to make the number divisible by 6. To a question such as "The sum of three successive numbers is ... What is the smallest number of these three?" (Item 25), 42 % of the black, 19 % of the Indian and 11 % of the white testees simply chose the distractor in which the smallest number appeared. Perhaps the reason for this shortcut to an answer was that many of the black pupils in the sample did not understand what was meant by "... sum of three successive numbers".

. Linear syllogisms. The only linear syllogism appearing in the JAT 2, namely Item 10, proved to be very easy for the white and Indian pupils, while the black testees found it very difficult. This is clear from the fact that 88 %, 71 % and 31 % of the testees respectively answered this item correctly. Because 29 % of the black testees chose Distractor E ("one cannot say"), it can be concluded that they either did not understand the question properly or did not try very hard to solve the problem. Another possible deduction that can be made is that a large percentage of these testees have not been attuned to critical thinking and they consequently display an inability to spot contradictions.

. Shuffled Letters. Items 24 and 27 of the JAT 2 consist of letters that have been shuffled and in order to answer the question, the testees have to rearrange these letters to form a word. From the responses it seems that a large percentage of the black pupils in particular answered the questions without reshuffling the letters. In the case of Item 24,

54 % of the black, 25 % of the Indian and 18 % of the white testees chose Distractor E, while 57 %, 22 % and 17 % of the testees respectively selected Distractor A as the answer to Item 27. The particular distractors would have been the correct answers if the testees had not been required to form a word from the shuffled letters. Two possible reasons can be put forward to explain why so many black testees experienced difficulties with these items: (1) They did not understand the instructions and therefore did not know exactly what to do. (2) They understood the instructions but were not able to reshuffle the letters to form a word. Language deficit apparently played a role in both cases.

To summarize: In the case of the Indian as well as the black testees bias in respect of verbal reasoning items can be ascribed to the functioning of a language factor. However language deficit is not the only reason a large percentage of the black testees did not stand the same chance as their white or Indian counterparts of answering an item correctly. The restricting factors do not so much emanate from a specific type of item, but they are largely the result of a particular form of test behaviour. Some of the most important issues in this regard are *inter alia* a tendency towards associative conceptualization, paying attention to the facts in an item in a selective manner, a carelessness concerning discrepancies and an uncritical attitude. The latter three point to lack of a logical attitude.

7.3.3 JAT 3: Number Ability

According to the mean difficulty values and discrimination values (Table A3 in Appendix A) the items functioned satisfactorily for the Indian testees and, moreover, the bias that did occur in a few odd cases (Table B3 in Appendix B) was in their favour. In the latter case Indian pupils performed better than their white counterparts of the same ability level (as determined by means of the Raven SPM).

The position differed somewhat as far as the black testees were concerned. Despite the fact that the discrimination values of the items were generally satisfactory ($r_{it} = \pm 0,3$), there were too many difficult items. It seems from Table A3 that the difficulty values of 16 of the 30 items were 0,4 and lower, while only four such items were encountered in the case of the white and five in the case of the Indian testees. From a comparison drawn between the easiest items (for example Items 1, 2, 3, 5, 7, 8 and 14) and the most difficult ones (for example Items 15, 16, 22, 26, 27 and 30), it appears that the latter group of items generally consisted of two-digit numbers that either had to be multiplied with or divided by a two-digit number while the former group largely consisted of a two-digit number (often smaller than 20) that had to be multiplied with or divided by a single-digit number. Thus it is clear that what played a role here was not so much bias - despite the significant indices in Table B3 - but rather a lack of skill in the handling of more difficult numbers. In some cases the testees also lacked the necessary knowledge. In the case of Item 16, for example, only 27 % of the black, compared with 50 % of the white and 44 % of the Indian testees, knew how to solve a problem such as $\frac{1}{7} \times 7 + 5$.

On the whole it is clear that this test is not suitable for the black pupils.

7.3.4 JAT 4: Synonyms

The JAT test that resulted in the most problems for the black pupils was Synomyms. According to Table A4 in Appendix A the mean difficulty values of the items were only 0,25 for these pupils (for the white and Indian testees these values were 0,67 and 0,52 respectively). The responses to the item distractors (Table A4) give an indication of the problems experienced by the black testees. It seems that in addition to displaying a lack of vocabulary, a large percentage of the pupils tended to form an association with the stimulus word rather than provided a synonym for it. Item 3 was a good example in this regard. Only 32 % of the testees indicated "warmth" to be the equivalent for "heat"; for 20,7 % "heat" was synonymous with "fire" and for 21 % it was synonymous with "burn".

Since it was the objective of this test to measure knowledge of vocabulary, the fact that the test as a whole was not suitable for the black testees cannot be attributed to bias as such, but rather to deficient knowledge of the language.

With regard to the Indian testees it can be deduced from the item statistics for difficulty and discrimination values that the test was suitable for them. The significant chi-square values in Table B4 in Appendix B should be attributed to the fact that the ability groups were formed on the basis of the scores obtained on the Raven SPM and not the scores obtained in the JAT 4 (the two tests have very little in common).

7.3.5 JAT 5: Comparison

Since this subtest consists of reasonably homogeneous items and aims at measuring perceptual speed, no item analysis was made, nor were bias indices calculated in this case. From the means reported in Table 4.1 it seems that the test functioned well both for Indian and black testees.

7.3.6 JAT 6 :(2-D) and JAT 7: (3-D)

Owing to the nature of the spatial tests not many of the items were expected to be biased. For this reason it was decided not to base the construction of the ability groups for calculating the chi-square indices on the total score of the SPM (as done in the case of other tests), but to use the total score of the specific test instead.

Homogeneous ability groups were formed on the basis of the following divisions of the test total score:
JAT 6: 00-14, 15-20, 21-28, 29-40 and
JAT 7: 00-10, 11-16, 17-21, 22-30.

JAT 6: 2-D
According to the data in Table 4.1 the test mean of the black pupils in the sample was two standard deviation units lower than that of the white testees, while the mean of the Indian pupils was one and a half standard deviation units lower than the test mean of their white counterparts.

Although the discrimination values of the items were generally satisfactory (cf. Table A5 in Appendix A), there were clearly too many items that the black testees found to be difficult. For them 20 of the 40

items had a difficulty value of 0,4 and lower, while for the Indian pupils there were 10 and for the white pupils only one such item.

Since reasonably few items with an index of bias that was significant on the 1 % level occurred among the black testees (Table B5) - the Scheuneman technique identified only two and the full chi-square technique eleven such items - the main reason for the poorer performance of these pupils cannot be ascribed to bias. In the case of the Indian testees the two methods identified five and eleven biased items respectively. There are no obvious differences between the biased and unbiased items: all the items consist of two figures that have to be combined to form a square. The fact that Item 40 proved to be biased against black as well as Indian testees and presupposed fine visual discrimination ability with regard to the relative lengths of vertical and horizontal lines nevertheless provided an indication of the possible nature of the problems experienced by some of the pupils. White testees also found this item very difficult (p = 0,39), but unlike in the case of the black testees this item still distinguished effectively between good and poor testees.

JAT 7: 3-D

According to the results in Table A6 (Appendix A) the difficulty values of 19 of the 30 items varied between 0,0 and 0,4 in respect of the black pupils in the sample. In the case of the Indian testees eight items fell in this category and in the case of the white testees not one item. Regarding the easier items, i.e. items with a p value of 0,6 and higher, only four such items were identified in the case of the black testees, seven in the case of the Indian and 25 in the case of the white testees. The discrimination values of virtually all these items were satisfactory (r_{it}= ± 0,4).

As in the case of the JAT 6, the number of biased items (Table B6 in Appendix B) does not offer a satisfactory explanation for the poorer performance of the Indian and particularly the black pupils in the sample. With regard to the black testees the Scheuneman technique identified two biased items at the one percent confidence level (Numbers 11 and 19) while seven such items were identified by means of the full chi-square (Numbers 9, 10, 11, 15, 17, 18 and 19). It was interesting to note that four of these items (Numbers 15, 17, 18 and 19) came from Section B of the test. In this division, which contains only eight items, testees had to form a three-dimensional figure according to a two-dimensional plan. The remaining three items came from Section A, in which two figures had to be combined to form a cube. Finding the correct answer to each of these three items demanded fine discrimination ability with regard to line lengths. Although Section C (Items 20 to 30), in which two three-dimensional figures had to be combined in a certain way, did not yield any biased items, the black testees generally found these items difficult and the p values varied between 0,27 and 0,48.

As far as bias against the Indian testees is concerned, only one item (Number 6) was identified with the help of the Scheuneman technique and seven items (Numbers 6, 8, 9, 14, 20, 24 and 27) with the help of the full chi-square technique (p < 0,01). With the exception of Item 6 the indices of bias proved to be not very large.

7.3.7 JAT 8: MEMORY (PARAGRAPH) AND JAT 9: MEMORY (WORDS AND SYMBOLS)

Because these tests require the testee merely to reflect what he/she can remember, no item analysis was made nor were any indices of bias calculated for the separate items.

JAT 8: Memory (Paragraph)

According to Table 4.1 the black testees performed particularly poorly in this test and their mean score was approximately half of what the white testees obtained. This result was not totally unexpected, since the paragraphs to be memorized were in English and the results of the Synonyms test have already indicated that the black pupils experienced difficulty with English vocabulary. On account of this language deficiency it can be deduced that the test did not measure the true memory ability of the black testees and it can therefore be considered biased. In the case of the Indian testees the test seemed to function satisfactorily.

JAT 9: Memory (Words and Symbols)

In contrast to their performance regarding the JAT 8, black testees performed very well in the JAT 9 and consequently showed that they had good associative abilities. In the light of the fact that the test reliabilities were virtually the same for all three population groups (Table 6.1), it seems justified to conclude that the test is unbiased for black as well as Indian testees.

7.3.8 JAT 10: Mechanical Insight

According to the test means (Table 4.1), the black pupils in the sample found the Mechanical Insight test particularly difficult and their score was almost the same as could be obtained by chance. At item level the test functioned extremely unsatisfactorily for the black as well as the Indian pupils in the sample. The mean difficulty value was only 0,26 for the black and 0,37 for the Indian testees, while the discrimination values were also low (0,19 for black and 0,24 for Indian testees). It is also clear from Table B7 (Appendix B) that only a small number of items did not display a significant chi-square value. Thus, based on their reasoning ability as reflected by their scores on the Raven SPM, black and Indian testees did not have the same chance as their white counterparts of answering a specific item correctly. In the case of the black pupils this unequal chance can apparently be attributed firstly to language deficit and secondly to deficient knowledge and insight with regard to physical principles. The latter problem may well be applicable to the Indian pupils as well.

7.4 SUMMARY

A summary of the most important causes of true and pseudo (apparent) item bias is given in Table 7.3. It seems that in the case of the Indian pupils bias was restricted mainly to items requiring knowledge of vocabulary, while for the black testees it appeared that bias was caused by a variety of factors. The most important of these were language deficit, lack of a logical attitude and a tendency towards association formation instead of conceptual reasoning. The forgoing is manifested in the form of item distractors which do not appeal to the same degree to the white, Indian and black pupils.

TABLE 7.3
SUMMARY OF PROBABLE CAUSES OF ABERRANT TEST ITEMS

TEST	DESCRIPTION	COMMENTS	GROUP
JAT 1: Classification	Illustrations of objects	* Lack of knowledge as to what the illustrations represent * Choose inconsistent answers * Do not consider all the information given	Indian & black black black
JAT 2: Reasoning	(i) Word Classification	* Insufficient vocabulary * Uncritical attitude * Reason by way of association rather than conceptualization (which leads to apparent or false item bias)	Indian & black black black
	(ii) Number Series	* Use only part of information - base answer on last two numbers	black
	(iii) Verbal Analogies	* Insufficient vocabulary * Association and not conceptualization * Ignore first part of the analogy	Indian & black black black
	(iv) Arithmetical "story" problems	* Insufficient vocabulary * Inaccurate reading (lack of concentration/motivation?)	Indian & black Indian & black
	(v) Linear Syllogisms	* Uncritical attitude (inability to see contradictions)	black
	(vi) Shuffled Letters	* Do not follow instructions (letters are not re-arranged) * Insufficient vocabulary	black black
JAT 3: Number Ability	Computations	* Lack of knowledge and skills	black
JAT 4: Synonyms	Synonymous words	* Insufficient vocabulary * Form associations with stimulus word	Indian & black black
JAT 6: 2-D	Combination of two 2-D figures	* Lack of fine visual discrimination ability concerning the lengths of the lines	Indian & black
JAT 7: 3-D	Manipulation of 3-D figures	* The same problem as with the JAT 6	Indian & black
JAT 10: Mechanical Insight	Drawings of mechanical devices	* Lack of knowledge and insight regarding physical principles * Insufficient vocabulary/ knowledge of the language	Indian & black black

8.0 DISCUSSION AND RECOMMENDATIONS

The most important aim of the investigation was to determine

- the suitability of the Junior Aptitude Tests (JAT), which was compiled and standardized for white pupils, for use with Indian and black pupils;
- the extent to which the JAT was biased (with regard to predictive validity, construct validity and test items) in respect of Indian and black pupils, and
- the possible causes of such bias.

Furthermore it was considered important, if possible, to lay down certain guidelines to be used in the construction of common tests in future.

8.1 THE SUITABILITY OF THE JAT FOR INDIAN AND BLACK PUPILS

At the beginning of the investigation it was expected that white pupils would perform better than the Indian and considerably better than the black pupils in the sample. This assumption was based on the white testees' more favourable socio-economic position and the privileges resulting from better educational facilities, hobbies and activities stimulating cognitive development in general.

In view of typical findings concerning black-white test differences in the USA (for example Jensen 1980), an expected difference in the test perfomance of white and black testees of approximately one standard deviation would not be unrealistic at all. However the results of the investigation revealed far bigger differences between the population groups than was the case in the USA. In addition to the fact that the white and the black testees' scores differed by two standard deviations and even more in respect of eight of the ten tests of the JAT, the means and variances were in some cases (e.g. JAT 2, JAT 3, JAT 4, JAT 8 and JAT 10) so small that these tests did not discriminate between good and poor testees at all. On account of the limited range of the score distributions it is not even possible to calculate separate norms for the black pupils in these cases.

The Indian testees found the JAT considerably easier than their black counterparts and although their performance, barring a few exceptions, was approximately one standard deviation lower than that of the white testees, the test means were in most cases at least reasonably close to the theoretical mean of the specific test. If the sizes of the means, variances and test reliabilities are taken into account, it can be concluded that the majority of tests in the JAT are suitable for use by these pupils.

Although a variety of reasons may have caused the mean test performance of the groups to differ, it was clear from the investigation that testing time was not one of the contributing factors. As could be expected, the extended testing time led to an improvement in performance, but this applied to the same extent to white, Indian and black pupils. In this investigation the measurement of test performance as a function of a particular testing time was not devoid of problems and it was not nearly as accurate as could have been possible if the measurement had taken place under laboratory conditions. However, despite the obvious short-

comings in the method that was adopted, it can be accepted that this method was at least suitable to indicate certain trends. Consequently the results revealed that the testing time could safely be ignored as a possible cause of the differences between the groups.

With regard to language as a possible factor contributing to group differences, a number of investigations indicated that the language of the test may contribute to cultural bias (Schultz & Fortune 1981) and in many cases it is actually responsible for the gap between verbal and non-verbal test scores (Jensen 1980). Owing to the fact that as from Standard 3 the medium of instruction for black pupils in South Africa is English, language was not expected to play a significant role in test performance in this investigation. However the results showed that this assumption was completely wrong. In fact, language played such an important role and the knowledge of English of the majority of black testees was so poor that certain tests, for example the JAT 4 (Synonyms) and the JAT 8 (Memory: Paragraph), proved to be virtually unusable.

Although the language factor undoubtedly contributed to the poorer performance of the black testees in certain tests, all differences between the groups cannot be related to language differences. The fact that the performance of black testees was considerably poorer than that of the other two groups in the non-language tests is adequate evidence of the effect of factors other than language only.

8.2 BIAS: PREDICTIVE VALIDITY, CONSTRUCT VALIDITY AND TEST ITEMS

8.2.1 Three aspects of predictive validity, i.e. differential validity, single-group validity and the regression equation (slope, intercept and standard error of estimate) were examined for bias. The results with regard to the JAT 1 (predictor) and the Raven SPM (criterion) appear in Tables 5.1, 5.2 and 5.3.

Differential validity implies that the correlation between the test and criterion scores differs significantly for different groups. According to Hunter, Schmidt and Hunter (1979), as well as Jensen (1980), this was not true in the case of comparisons between white and black groups in the USA. The findings of this investigation correspond in great measure with their viewpoint. The correlation coefficients of white and black pupils did not differ significantly, and although a significant difference was found between the correlations of white and Indian testees, overlapping confidence intervals indicated that this difference was not substantial.

Single-group validity implies that the validity coefficients are significantly higher than zero for only one of the groups involved. The results of the investigation indicated that this form of bias did not occur in any of the groups. The results confirmed a trend often encountered in the USA (cf. Hunter & Schmidt 1978; Reynolds 1980).

Before it can be concluded that bias did not occur with regard to predictive validity, it has to be indicated that the slopes, intercepts and standard errors of estimate of the regression lines for the particular groups did not differ significantly. However significant differences regarding each of the three factors mentioned were found in this investigation (Table 5.2).

Another important question to be considered was the extent to which the Indian and black pupils would be at a disadvantage when the regression equation for white pupils was used to predict their criterion score. According to investigations in the USA the performance of minority groups was generally overpredicted when a common regression equation was used and these groups actually benefited from bias (Kallingal 1971; Temp 1971; Goldman & Richards 1974; Goldman & Hewitt 1976). The findings of this investigation correspond with those reached in the USA. According to Table 5.3 the criterion performance of the black pupils was overpredicted by 15 points on average(i.e. 1,2 of a standard deviation) when the regression equation for white pupils was used. Even when the correlation coefficient was corrected for attenuation owing to unreliabilities, the overprediction still remained approximately one standard deviation.

The relation between a predictor and a criterion displays certain shortcomings as far as the investigation of bias is concerned. Firstly the reliabilities of the two measurements are hardly ever identical for the two groups concerned and this may result in the predictor appearing biased while this is not really the case. Secondly, although the validity coefficients of two groups may be identical, the criterion score of one group may still be underpredicted or overpredicted if the regression equation of the other group is used. Consequently one can agree with Drasgow (1982) that, in statistical terms, a differential validity investigation is not a powerful approach to the detection of test bias.

The fact that differential and single-group validity - as an indication of bias - did not occur in the JAT 1 does not automatically imply that it will not occur in the other tests of the JAT either. Firstly, of all the tests in the JAT, the JAT 1 functioned best for both Indian and black testees. Furthermore the SPM is an ideal criterion for evaluating the predictive validity of the JAT 1 since both this test and the SPM are non-verbal tests aimed at measuring inductive reasoning. Thus, in agreement with Katzell and Dyer (1978) it is concluded that at this stage it would not be possible to ignore the question of differential and single-group validity as if it were non-existent.

8.2.2 Regarding bias in construct validity, that is, the extent to which a test battery measures the same psychological constructs in different groups, three of the most important criteria were examined with regard to the white, Indian and black pupils in the sample. These were

. similar test reliabilities
. similar rank order correlations of item difficulty values, and
. factorial similarity.

With regard to similar test reliabilities it appears from Table 6.1 that in most cases statistically significant differences can be ascribed to the large number of degrees of freedom. With the exception of the JAT 10 (Mechanical Insight) in respect of Indian testees and the JAT 2 (Reasoning), JAT 4 (Synonyms) and JAT 10 in respect of black testees, the test reliabilities of the three groups are virtually the same. The fact that language items were common to all the tests revealing low reliabilities once again focused the attention on the importance of this factor as a cause of the differences between the population groups.

Although the reliabilities are the same in most cases, this does not imply that the JAT measures the same constructs in the three groups. Test reliability is at any rate a poor criterion of bias in construct validity because the same test can measure different constructs in different groups with the same degree of accuracy. However similar reliabilities are a necessary prerequisite for non-bias, and with the exception of the tests mentioned above, most of the JAT tests meet this requirement.

Next the similarity of the rank order correlations of the item difficulty values was examined for the three groups (a comparison was drawn between the rank order of the item difficulty values of the Indian and the black testees on the one hand and that of the white testees on the other hand). When the same construct is measured in the three groups, the rank order of the item difficulty values should be more or less the same for the groups. The results in Table 6.2 show that, with the exception of the JAT 3 (Number Ability), the difficulty values of the items were arranged differently for white and black testees. Based on the strong interaction found between population group and item difficulty values in respect of the JAT 4 (Synonyms) and JAT 10 (Mechanical Insight) for the black pupils and the JAT 4 for the Indian pupils, it can furthermore be deduced that language was at least partially responsible for the results. However it is clear that other factors also played a role, judging by the strong interaction revealed by a non-verbal test such as the JAT 1 (Classification). More detailed attention will be given to these factors in the discussion of item bias.

8.2.3 In contrast to the previous two methods, factor analysis gives a more precise indication not only of the constructs measured, but also of the extent to which the same constructs are measured in the different groups. In the case of white testees three factors were found by means of exploratory factor analysis. These three factors were identified as a spatial, memory (including other elements) and verbal factor respectively (Table 6.5). These factors should nevertheless be regarded as merely tentative since they are defined by only two instead of three or more variables. The specific "name" of a factor is not very important within the context of bias in construct validity; what is important, however, is whether the same factors occur in the various groups.

Two additional analyses, namely coefficients of congruence (Table 6.6) and the method of rotation to a least squares fit given a partially specified target (Table 6.7) showed that the results did not contradict the hypothesis that the JAT measures the same constructs among the white, Indian and black pupils in the sample. The dimensions underlying the abilities as measured by the JAT are to a large extent similar for the three population groups. This finding has important implications for the construction of common tests for different population groups: Despite great differences, particularly between white and black testees as far as the mean level of ability (as displayed in Table 4.1) is concerned, the differences are still not so fundamental that different psychological principles are required to explain the behaviour of the two groups. This also implies that differences in test performance as such do not necessarily affect the construct validity of a test.

The main finding of this investigation, namely that population group affects the difficulty value of ability tests but not necessarily the factor pattern, corresponds with the results of several researchers abroad, for example Stodolsky and Lesser (1967), Miele (1979), Sandoval (1979),

Sung and Dawis (1981), Reynolds (1983) and Mercer (1984). The results of the investigation also support the viewpoints of Hakstian and Vandenberg (1979) and of Van der Flier (1980). According to them there is greater correspondence in the cognitive structure of different cultural groups than is generally believed.

Based on findings similar to those reached in the current investigation (for example the absence of bias in construct validity), Jensen (1980) and Reynolds (1980) concluded that the differences in mean test performance between whites and blacks in the USA signified actual differences regarding the ability of these groups. However this conclusion cannot simply be generalized to the population groups in South Africa. For example the results of the current investigation clearly show that language played a prominent role in all the tests containing language items and this had a negative effect on the performance of the black testees in particular. The findings also revealed that the absence of bias in construct validity did not exclude the possibility of biased items systematically underestimating the ability of a group, either as the result of language or other factors. When language as such is measured, as in the case of the JAT 4 (Synonyms), the discrepancy between the groups probably signifies an actual difference in respect of the specific ability. In this case the discrepancy is not necessarily a result of bias. An alternative explanation would imply that all spelling tests are biased in the case of poor spellers or that all arithmetic tests are biased in the case of those persons who are not able to do additions or subtractions.

8.2.4 Regarding item bias it is clear that in addition to language and general knowledge, which undoubtedly make an important contribution to differential performance, there must have been other factors that also resulted in the ability of the black pupils being underestimated.

The crux of the problem is summarized neatly by Ironson and Subkoviak (1979): *One pervasive question in bias research is whether differences are due to ability or to bias* (pp. 213-214).

On account of certain idiosyncratic forms of test behaviour revealed by a large percentage of black testees, it was very difficult to determine whether poor performance could be attributed to bias or inability. For example testees either did not restrict themselves to the information given in an item, or used only part of the information. In addition, testees not only selected distractors that represented the poorest possible choice from the five options, but also distractors that were often directly in contrast with the stem of the item, etc.. Such results are characteristic of persons with a low SES, inadequate schooling and deficient test sophistication. Despite the fact that it would be possible to explain these forms of test behaviour as a result of cultural lag, it remains insufficient reason to classify an item that evokes this type of behaviour *ipso facto* as biased.

On the basis of the test behaviour of the black pupils certain important "shortcomings" (that is, from the point of view of a test constructor in a Western situation) could be perceived in the strategies that the testees adopted to solve problems. The first such problem that will be discussed occurred in classification-type items, in other words the categorization of objects. Categorization is a basic cognitive process found in most intellectual tasks and cross-cultural differences in this

regard point to a tendency to classify objects according to thematical rather than taxonomical criteria (Markman & Callanan 1984). The findings of the present investigation strongly supported this view, but the term "associative thought or reasoning" was used instead of "thematical", since in the author's opinion the former was a better description of the particular cognitive process. The problem-solving strategy can be illustrated with the aid of an imaginary item in the following manner:

"What fits best with KNIFE and FORK?"

A. SPOON B. PLATE C. TABLE

Most white pupils would choose A since the three objects are conceptually similar; this is a typical left-brain response (cf. Kaufman 1979). Most black pupils, however, would select B as a result of the association formed with regard to KNIFE and FORK; this manner of reasoning, which is basically holistic, is a typical right-brain response (cf. Kaufman 1979). The possibility that preference for a taxonomical selection largely depends on schooling (Markman & Callanan 1984) has important implications for the construction of a commom test for different population groups.

A second category of items reflecting specific peculiarities in the test behaviour of the black testees is analogies. In order to identify problems experienced by a large percentage of these pupils, it is essential to concentrate briefly on the processes underlying an analogy item. A typical analogy item represents a proportional format, i.e. A : B :: C ?, and according to Pellegrino (1985) the process according to which such an item is answered can be divided into three general classes. The first of these classes is the discovery of the relevant characteristics or encoding, during which the most important characteristics of each term in the analogy are represented in memory. In the case of verbal items this process comprises the activation of a number of semantic characteristics associated with the concept, while in the case of figure items encoding requires a description of individual forms or elements, as well as of their specific characteristics and relations. The discovery of characteristics is an essential prerequisite for the execution of the subsequent processes. The second class comprises those processes that are used for comparing the characteristics of specific pairs of terms. The first such process - inference - concerns the relation between the first two terms of the analogy (A : B). In the case of verbal items the inference process serves to discover the semantic characteristics linking the two concepts directly or indirectly. It also entails the application of a specific rule, which has been deduced in respect of the A : B pair, to the characteristics of the C term in order to establish an ideal solution for the D term. The third class of processes evaluates the applicability of the solution proposed for the D term.

When the results of the investigation are interpreted in the light of the previously described processes, it is quite clear that a large percentage of the black pupils in the sample already experienced problems during the first phase, namely the discovery of characteristics or encoding phase. As this phase was crucial, the black pupils were consequently incapable of finding the correct solutions. With regard to verbal analogies one possible reason according to the findings of

Resnick and Neches(1984), and Sternberg(1986), for the failure of the encoding phase among the testees may be that they were not familiar with one or more of the words that made up the terms. The naivety of the testees regarding the handling of analogy-type problems might possibly have been an even more important cause in this regard.

Series Completion constitutes a third category of items that further illustrated the test behaviour of black testees. In order to put the findings of the investigation into proper perspective, attention should first of all be given to the processes underlying the selection of answers to the Series Completion items. In addition to processes that are basically the same as those used for analogies (Sternberg 1986), the problem-solving processes also reveal certain specific characteristics (Pellegrino 1985). The first of these components require the testee to scan the series and to formulate ideas on the possible relation between the elements. The second problem-solving component is the discovery of the periodicity of the series, in other words the number of elements that represent one complete cycle of the pattern. (The period length of a series such as 2, 5, 8, 11 ? is one, since the same relation is applicable to every subsequent element. The period length of a series such as 9, 9, 10, 10, 11, 11, ? is two, as two numbers occur every time before the cycle is repeated.) Finally the testee has to identify the relation controlling the positions within a period and deduce a general rule for the particular series. This rule should then be applied to find the missing number, letter or figure. It appeared from the investigation that a larger percentage of the black than of the Indian or white testees tried to find a rule by concentrating on the last two elements of a series only. The fact that some testees did not follow the correct procedure for problem-solving may be attributed either to an inability to spot relations or to a lack of knowledge about the handling of a series in general. According to the available data it was not possible to determine which one of these two probable reasons was the most important.

In view of the preceding discussion of the processes underlying the answering of certain types of items, as well as the results of this investigation concerning the responses of the testees, the following general principle can be formulated to explain deviating items, i.e. items that do not fit in with the general pattern of items in the test: Abnormality in item behaviour may result from true bias, apparent bias or of a combination of these two forms of bias. The reason why previous researchers have so far not been able to find a significant number of generalizable principles regarding item bias may on the one hand be attributed to the fact that fewer differences are usually encountered between groups than is often suspected - at least as far as the USA is concerned - and, on the other hand, to the fact that a clear distinction is not drawn between true and apparent item bias.

Aspects of true item bias that were revealed by this investigation include the following:

- Language factors, i.e. the use of words and concepts that are not understood in items in which language should be irrelevant
- Distractors that are too tempting and confuse those testees who are less test-sophisticated

- Knowledge and concepts that are obtained mainly in an informal manner and in respect of which there is proof that the general knowledge basis of the groups is not comparable.

Apparent item bias mainly concerns the test behaviour of the testees and therefore the cause of the bias is not found within the scope of the item itself. The investigation revealed the following causes:

- Associative conceptualization
- Selective attention to the information contained in an item
- Deviating from the facts in an item
- Lack of a logical attitude
- Lack of specific knowledge that is usually obtained in a formal manner, in other words at school.

Attention will have to be given to both forms of item bias mentioned above if a common test that even distantly meets the requirements of non-bias is to be constructed for the different population groups in South Africa.

8.3 RECOMMENDATIONS

In order to render it possible to compile a common test battery for all the different population groups in South Africa, and for white and black pupils in particular, it is first of all essential that the gap between the two groups mentioned be narrowed as far as their mean test performance is concerned. This implies that factors leading in an irrelevant manner to differential test performance should be reduced or eliminated as far as possible.

One such factor that was highlighted by the investigation was that the white and the black pupils in the sample did not find the item distractors equally appealing. Certain distractors that most white testees regarded as completely wrong apparently had strong persuasive power for many of the black pupils. Test constructors usually strive to include distractors that are as tempting as possible in order to distract the attention of those testees who do not really know the correct answer. In their attempt to put such pupils on a wrong track, the typical thinking errors that are usually made by testees are imitated. Despite the fact that it is standard practice to use this method for constructing items, it does not only lead to confusion among the less sophisticated testees because they are not able to recognize the correct answer, but also reinforces certain undesirable thinking patterns.

It is consequently recommended that test constructors once again consider the procedures and strategies on which the construction of test items is based and that special attention be given to item distractors. It is of course impossible to try and prevent pupils from reasoning illogically by means of the distractors, but by not providing a distractor that for example represents an associative solution, the testees can be forced at least to consider other possible options as well. In the case of unsophisticated testees attempts to mislead by means of distractors should generally be avoided as far as possible.

The investigation showed that a large percentage of the black pupils in the sample did not have the knowledge or skills to exploit the properties and format of the different item types in such a way as to obtain

high test scores. Illustrating the principles underlying the responses to the different item types with the aid of a limited number of practice examples was apparently not sufficient to provide these testees with the necessary insight and comprehension. Based on the views of Alley and Foster (1978), Borkowski and Krause (1983), Frederiksen (1984) and Van de Vijver, Daal and Zonneveld (1986) concerning practice and training, it is recommended that urgent attention be given to the training of language and cognitive skills in general and problem-solving strategies in particular. In order to render the test performance of white and black testees comparable, black pupils should be equipped with a basic "supply" of reasoning strategies on which they can draw to obtain solutions to problems encountered in item types such as analogies, series completion, classification and statements. It is consequently recommended that training tests be developed which illustrate the principles required for answering the different item types often encountered in ability tests. No testee should sit for the real test unless he can prove that he has mastered certain basic problem-solving skills. Of course this is just a short-term solution to a problem that lies much deeper and involves far more than a mere lack of test sophistication.

As far as a solution in the long term is concerned, it is recommended that experimental work be conducted on the item formats and item contents leading to the differential performance of groups.

The macro-approach adopted with regard to the study of item bias in this investigation has limited the amount of information that can be obtained from the responses given by the testees. A micro-approach in the tradition of cognitive psychologists such as Sternberg, Pellegrino, Glaser, Hunt and others would render it possible to study the cognitive processes underlying the answering of items and offer an explanation for the apparently illogical test behaviour of some testees.

In conclusion it can be stated that the construction of common tests for the different population groups in South Africa is one of the major challenges to be met before the turn of the century. If, after all that has so far been done in the USA in this regard, Scheuneman (1985) can still remark "*...how little we know about the mechanisms that produce differential performance between Blacks and Whites...*" (p. 49), one can appreciate the task that lies ahead.

REFERENCES

Alley, G. & Foster, C. 1978. Nondiscriminatory testing of minority and exceptional children. Focus on Exceptional Children, 9, 1-14.

Anastasi, A. 1976. Psychological testing. New York: MacMillan, Fourth Edition.

Anastasi, A. 1985. Some emerging trends in psychological measurement: A fifty-year perspective. Applied Psychological Measurement, 9(2), 121-138.

Baker, F.B. 1981. A criticism of Scheuneman's item bias technique. J. of Educational Measurement, 18(1), 59-62.

Bernal, E.M. 1975. A response to "Educational uses of tests with disadvantaged students." American Psychologist, 30, 93-95.

Boehm, V.R. 1978. Populations, preselection, and practicalities: a reply to Hunter and Schmidt. J. of Applied Psychology, 63(1), 15-18.

Bond, L. 1981. Bias in mental tests. New Directions for Testing and Measurement, 11, 55-77.

Borkowski, J.G. & Krause, A. 1983. Racial differences in intelligence: The importance of the executive system. Intelligence, 7, 379-395.

Browne, M.W. 1972. Orthogonal rotation to a partially specified target. British J. of Mathematical and Statistical Psychology, 25, 115-120.

Browne, M.W., Schoeman, A. & Glas, A.P. 1979. 'n Program vir die gelyktydige rotasie van verskeie faktormatrikse na maksimum ooreenkoms en eenvoudige struktuur. (A programme for the simultaneous rotation of various factor matrices to maximum similarity and simple structure.) Pretoria: Human Sciences Research Council.(Unpublished document)

Burrill, L. E. 1981. A comparative investigation into the identification of ethnic bias in items assessing current educational status. Unpublished PH.D. Dissertation, Fordham University, New York. (Univ. Microfilms International, Ann Arbor, Michigan, 1984)

Burrill, L.E. & Wilson, R. 1980. Fairness and the matter of bias. Test Service Notebook 36. New York: The Psychological Corporation.

Carlson, J.S. & Jensen, C.M. 1981. Reliability of the Raven Colored Progressive Matrices Test: Age and ethnic group comparisons. J. of Consulting and Clinical Psychology, 49(3) 320-322.

Carter, J.E. & Wilkinson, L. 1984. A latent trait analysis of the MMPI. Multivariate Behavioral Research, 19, 385-407.

Cattell, R.B. 1940. A culture-free intelligence test 1. J. of Educational Psychology, 31. 161-179.

Cleary, T.A. 1968. Test bias: prediction of grades of Negro and White students in integrated colleges. J. of Educational Measurement, 5(2), 115-124.

Cleary, T.A. & Hilton, T.L. 1968. An investigation of item bias. Educational and Psychological Measurement, 28, 61-75.

Cleary, T.A., Humphreys, L.G., Kendrick, S.A. & Wesman, A. 1975. Educational uses of tests with disadvantaged students. American Psychologist, 30, 15-41.

Cole, N.S. 1981. Bias in testing. American Psychologist, 36(10), 1067-1077.

Directorate: Liaison and Information 1985. Reaction of the Government on the HSRC report on intergroup relations. Pretoria: State President's Office. (Media release by State President P.W. Botha)

Divgi, D.R. 1985. A minimum chi-square method for developing a common metric in item response theory. Applied Psychological Measurement, 9(4), 413-415.

Dorans, N.J. 1979. The need for a common metric in item bias studies. TM 79-20. Washington, D.C. : U.S. Office of Personnel Management.

Drasgow, F. 1982. Biased test items and differential validity. Psychological Bulletin, 92(2), 526-531.

Durovic, J.J. 1975. Test bias: An objective definition for test items. Paper presented at the Annual Meeting of the Northeastern Educational Research Association, Ellenville, New York. (ERIC Document Reproduction Service No. ED 128 381)

Evans, F.R. & Reilly, R.R. 1972. A study of test speededness as a potential source of bias in the admission test for Graduate Study in Business quantitative score. Report RB-72-9. Princeton, N.J.: Educational Testing Service.

Flaugher, R.L. 1978. The many definitions of test bias. American Psychologist, 33(7), 671-679.

Frederiksen, N. 1984. Implications of cognitive theory for instruction in problem solving. Review of Educational Research, 54(3), 363-407.

Goldman, R.D. & Hewitt, B.N. 1975. An investigation of test bias for Mexican-American college students. J. of Educational Measurement, 12(3), 187-196.

Goldman, R.D. & Hewitt, B.N. 1976. Predicting the success of Black, Chicano, Oriental and White college students. J. of Educational Measurement, 13(2), 107-117.

Goldman, R.D. & Richards, R. 1974. The SAT prediction of grades for Mexican-American versus Anglo-American students at the University of California, Riverside. J. of Educational Measurement, 11(2), 129-135.

Green, D.R. 1972. Racial and ethnic bias in test construction. Monterey, California: CTB/McGraw-Hill.

Hakstian, A.R. & Vandenberg, S.G. 1979. The cross-cultural generalizability of a higher-order cognitive structure model. Intelligence, 3(1), 73-103.

Hambleton, R.K. 1980. Latent ability scales: Interpretations and uses. New Directions for Testing and Measurement, 6, 73-97.

Hambleton, R.K. & Cook, L.L. 1977. Latent trait models and their use in the analysis of educational test data. J.of Educational Measurement, 14(2), 75-96.

Hambleton, R.K., Swaminathan, H., Cook, L.L., Eignor, D.R. & Gifford, J.A. 1978. Developments in latent trait theory: Models, technical issues, and applications. Review of Educational Research, 48(4), 467-510.

Manual (1980-edition) for the Junior Aptitude Tests (JAT) 1980. Pretoria: Human Sciences Research Council.

Harman, H.H. 1967. Modern factor analysis.(2nd ed.) Chicago: University of Chicago Press.

Hui, C.H. & Triandis, H.C. 1985. Measurement in cross-cultural psychology. J. of Cross-Cultural Psychology, 16(2), 131-152.

Hulin, C.L., Drasgow, F. & Parsons, C.K. 1983. Item response theory: Application to psychological measurement. Homewood, Illinois: Dow Jones-Irwin.

Humphreys, L.G. 1986. An analysis and evaluation of test and item bias in the prediction context. J.of Applied Psychology, 71(2), 327-333.

Humphreys, L.G. & Taber, T. 1973. Ability factors as a function of advantaged and disadvantaged groups. J. of Educational Measurement, 10(2), 107-115.

Hunter, J.E. & Schmidt, F.L. 1978. Differential and single-group validity of employment tests by race: A critical analysis of three recent studies. J. of Applied Psychology, 63(1), 1-11.

Hunter, J.E., Schmidt, F.L. & Hunter, R. 1979. Differential validity of employment tests by race: A comprehensive review and analysis. Psychological Bulletin, 86(4), 721-735.

Hunter, J.E., Schmidt, F.L. & Rauschenberger, J. 1984. Methodological, statistical, and ethical issues in the study of bias in psychological tests. In C.R. Reynolds & R.T. Brown(Eds.), Perspectives on bias in mental testing. New York: Plenum.

Intasuwan, P. 1979. A comparison of three approaches for determining item bias in cross- national testing. Unpublished D.Phil. Dissertation, University of Pittsburgh. (University Microfilms International, Ann Abor, Michigan, 1982)

Ironson, G.H. & Subkoviak, M.J. 1979. A comparison of several methods of assessing item bias. J. of Educational Measurement, 16(4), 209-225.

Jensen, A.R. 1980. Bias in mental testing. London: Methuen.

Kallingal, A. 1971. The prediction of grades for black and white students at Michigan State University. J. of Educational Measurement, 8(4), 263-265.

Katzell, R.A. & Dyer, F.J. 1978. On differential validity and bias. J. of Applied Psychology, 63(1), 19-21.

Kaufman, A.S. 1979. Cerebral specialization and intelligence testing. J. of Research and Development in Education, 12(2), 96-107.

Kelley, H.P. 1982. Are culturally biased tests useful? New Directions for Testing and Measurement, 16, 125-133.

Kim, J.- O. & Mueller, C.W. 1978. Factor analysis: Statistical methods and practical issues. Beverly Hills: Sage Publications.

Kulick, E. & Dorans, N.J. 1984. The standardization approach to assessing unexpected differential item performance. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, April 23-27. (ERIC Document Reproduction Service No. ED 247 260)

Lawley, D.N. & Maxwell, A.E. 1963. Factor analysis as a statistical method. London: Butterworths.

Linn, R.L., Levine, M.V., Hastings, C.N. & Wardrop, J.L. 1980. An investigation of item bias in a test of reading comprehension. Technical Report No. 163. Center for the study of reading. Champaign, Illinois: University of Illinois at Urbana-Champaign. (ERIC Document Reproduction Service No. ED 184 091)

Linn, R.L., Levine, M.V., Hastings, C.N. & Wardrop, J.L. 1981. Item bias in a test of reading comprehension. Applied Psychological Measurement, 5(2), 159-173.

Lord, F.M. 1980. Application of item response theory to practical testing problems. Hillsdale, N.J.: Lawrence Erlbaum.

Lord, F.M. & Wingersky, M.S. 1984. Comparison of IRT true-score equipercentile observed-score "equatings". Applied Psychological Measurement, 8(4), 453-461.

Loyd, B. 1983. The effect of number of ability intervals on the stability of item bias detection. Paper presented at the Annual Meeting of the Eastern Educational Research Association. Baltimore, February 24-27. (ERIC Document Reproduction Service No. ED 247 297)

Marascuilo, L.A. & Slaughter, R.E. 1981. Statistical procedures for identifying possible sources of item bias based on χ^2 statistics. J. of Educational Measurement, 18(4), 229-248.

Markman, E.M. & Callanan, M.A. 1984. An analysis of hierarchical classification. In R.J. Sternberg(Ed.), Advances in the psychology of human intelligence. Volume 2. Hillsdale, N.J.: Lawrence Erlbaum.

McKenna, F.P. 1984. Measures of field dependence: Cognitive style or cognitive ability? J. of Personality and Social Psychology, 47(3), 593-603.

Mellenbergh, G.J. 1983. Conditional item bias methods. In S.H. Irvine & J.W. Berry(Eds.), Human assessment and cultural factors. New York: Plenum Press.

Messick, S. 1980. Test validity and the ethics of assessment. American Psychologist, 35(11), 1012-1027.

Miele, F. 1979. Cultural bias in the WISC. Intelligence, 3(2), 149-164.

Osterlind, S.J. 1983. Test item bias. Beverly Hills: Sage Publications.

Pellegrino, J.W. 1985. Inductive reasoning ability. In R.J. Sternberg (Ed.), Human abilities: An information-processing approach. New York: W.H. Freeman.

Petersen, N.S. 1980. Bias in the selection rule- bias in the test. In L.J.T. van der Kamp, W.F. Langerak & D.N.M. de Gruijter (Eds.), Psychometrics for educational debates. Chichester: John Wiley.

Powers, S. & Jones, P.B. 1984. Factorial invariance of the California Achievement Tests across race and sex. Educational and Psychological Measurement, 44, 967-970.

Raju, N.S. & Normand, J. 1985. The regression bias method: A unified approach for detecting item bias and selection bias. Educational and Psychological Measurement, 45, 37-54.

Raven, J.C., Court, J.H. & Raven, J. 1977. Manual for Raven's Progressive Matices and Vocabulary scales. London: H.K. Lewis & Co.

Ree, M.J. 1979. Estimating item characteristic curves. Applied Psychological Measurement, 3(3), 371-385.

Resnick, L.B. & Neches, R. 1984. Factors affecting individual differences in learning ability. In R.J. Sternberg(Ed.), Advances in the psychology of human intelligence. Volume 2. Hillsdale, N.J.: Lawrence Erlbaum.

Reynolds, C.R. 1980. Differential construct validity of intelligence as popularly measured: Correlations of age with raw scores on the WISC-R for blacks, whites, males and females. Intelligence, 4(4), 371-379.

Reynolds, C.R. 1982. Methods for detecting construct and predictive bias. In R.A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Reynolds, C.R. 1983. Test bias: In God we trust: all others must have data. The J. of Special Education, 17(3), 241-260.

Reynolds, C.R. & Brown, R.T. 1984. Bias in mental testing: An introduction to the issues. In C.R. Reynolds & R.T. Brown(Eds.), Perspectives on bias in mental testing. New York: Plenum Press.

Roos, W.L. 1984. Prestasieprofiele vir hoëvlakwerkkrag. (Achievement profiles for high-level manpower.) Report MT-53. Pretoria: Human Sciences Research Council.

Rudner, L.M. & Convey, J.J. 1978. An evaluation of select approaches for biased item identification. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada, March 27-31. (ERIC Document Reproduction Service No. ED 157 942)

Rudner, L.M., Getson, P.R. & Knight D.L. 1980. Biased item detection techniques. J. of Educational Statistics, 5(2), 213-233.

Rudner, L.M. & Getson, P.R. 1982. Item bias research has its limitations. Paper presented at the Annual Meeting of the National Council for Measurement in Education, New York, March.

Sandoval, J. 1979. The WISC-R and internal evidence of test bias with minority groups. J. of Consulting and Clinical Psychology, 47(5), 919-927.

Scheuneman, J.D. 1979. A new method of assessing bias in test items. J. of Educational Measurement, 16(3), 143-152.

Scheuneman, J.D. 1980. Latent-trait theory and item bias. In L.J.T. van der Kamp, W.F. Langerak & D.N.M. de Gruijter(Eds.), Psychometrics for educational debates. Chichester: John Wiley.

Scheuneman, J.D. 1981a. A new look at bias in aptitude tests. New Directions for Testing and Measurement, 12, 3-35.

Scheuneman, J.D. 1981b. A response to Baker's criticism. J. of Educational Measurement, 18(1), 63-66.

Scheuneman, J.D. 1985. Exploration of causes of bias in test items. Research Report 85-42. Princeton, N.J.: Educational Testing Service.

Schmidt, F.L., Pearlman, K. & Hunter, J.E. 1980. The validity and fairness of employment and educational tests for Hispanic Americans: A review and analysis. Personnel Psychology, 33(4), 705-724.

Schultz, L.J. & Fortune, J.C. 1981. The three I's: Sources of test bias. Education, 102(2), 117-123.

Shepard, L.A. 1981. Identifying bias in test items. New Directions for Testing and Measurement, 11, 79-104.

Shepard, L.A. 1982. Definitions of bias. In R.A. Berk(Ed.), Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

Shepard, L.A., Camilli, G. & Averill, M. 1981. Comparison of procedures for detecting test-item bias with both internal and external ability criteria. J. of Educational Statistics, 6(4), 317-375.

Shepard, L.A., Camilli, G & Williams, D.M. 1983. Accounting for statistical artifacts in item bias research. Paper presented at the Annual

Meeting of the American Educational Research Association, Montreal, April.

Sternberg, R.J. 1986. Toward a unified theory of human reasoning. Intelligence, 10(4), 281-314.

Stocking, M.L. & Lord, F.M. 1983. Developing a common metric in item response theory. Applied Psychological Measurement, 7(2), 201-210.

Stodolsky, S.S. & Lesser, G. 1967. Learning patterns in the disadvantaged. Harvard Educational Review, 37, 546-593.

Subkoviak, M.J., Mack, J.S., Ironson, G.H. & Craig, R.D. 1984. Empirical comparison of selected item bias detection procedures with bias manipulation. J. of Educational Measurement, 21(1), 49-58.

Sundberg, N. D. & Gonzales, L.R. 1981. Cross-cultural and cross- ethnic assessment: Overview and issues. In P. McReynolds(Ed.), Advances in psychological assessment. Volume 5. San Francisco: Jossey-Bass.

Sung, Y.H. & Dawis, R.V. 1981. Level and factor structure differences in selected abilities across race and sex groups. J. of Applied Psychology, 66(5), 613-624.

Temp, G. 1971. Validity of the SAT for Blacks and Whites in thirteen integrated institutions. J. of Educational Measurement, 8(4), 245-251.

Thorndike, R.L. & Hagen, E. 1969. Measurement and evaluation in psychology and education. New York: John Wiley.

Traub, R.E. & Lam, Y.R. 1985. Latent structure and item sampling models for testing. Annual Review of Psychology, 36, 19-48.

Van de Vijver, F.J.R., Daal, M. & Van Zonneveld, R. 1986. The trainability of formal thinking: A cross-cultural comparison. International J. of Psychology, 21, 589-615.

Van der Flier, H. 1980. Vergelijkbaarheid van individuele testprestaties. Lisse: Swets & Zeitlinger.

Van der Flier, H. & Drenth, P.J.D. 1980. Fair selection and comparability of test scores. In L.J.T. van der Kamp, W.F. Langerak & D.N.M.de Gruijter(Eds.), Psychometrics for educational debates. Chichester: John Wiley.

Warm, T.A. 1978. A Primer of item response theory. Technical report 941078. Oklahoma City: U.S. Coast Guard Institute.

Weiss, D.J. & Davidson, M.L. 1981. Test theory and methods. Annual Review of Psychology, 32, 647-651.

Wood, R.L., Wingersky, M.S. & Lord, F.M. 1976. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters. Research Memorandum RM -76-6 (modified 1/78). Princeton, N.J. : Educational Testing Service.

APPENDIX A

Percentage of testees who selected a particular option, proportion correct(p) and item discrimination values (r_{it}) for the items of seven tests of the JAT

TABLE A1
JAT 1: Classification

Item	White							Indian							Black						
	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}
	1	2	3	4	5			1	2	3	4	5			1	2	3	4	5		
1	7,9	3,3	0,3	0,1	88,4*	0,88	0,25	8,7	5,7	1,3	0,3	84,0*	0,84	0,29	3,6	5,8	1,4	0,6	88,6*	0,89	0,21
2	0,4	0,2	3,3	96,1*	0,0	0,96	0,24	1,6	2,4	4,0	91,8*	0,2	0,92	0,25	8,8	7,1	6,5	77,0*	0,6	0,77	0,36
3	0,9	0,7	92,5*	4,2	1,7	0,93	0,18	0,9	3,0	83,9*	8,0	4,3	0,84	0,29	0,7	1,5	83,4*	12,3	2,1	0,83	0,13
4	0,4	95,6*	1,8	1,2	0,9	0,96	0,23	1,2	94,6*	3,1	0,3	0,8	0,95	0,25	1,1	84,3*	12,8	0,7	1,1	0,84	0,32
5	1,3	3,1	2,5	90,0*	3,0	0,90	0,27	2,1	3,6	7,6	79,9*	6,8	0,80	0,36	4,4	7,8	17,9	59,3*	10,5	0,59	0,33
6	0,9	82,2*	1,3	6,4	9,2	0.82	0,28	0,6	84,8*	1,8	4,8	8,0	0.85	0.19	2,7	65,6*	2,5	15,7	13,5	0,66	0,28
7	5,8	1,7	3,6	1,2	87,6*	0,88	0,20	5,5	3,7	5,3	1,2	84,4*	0,84	0,27	4,1	8,0	5,4	1,7	80,9*	0,81	0,26
8	4,7	85,1*	6,6	0,3	3,2	0,85	0,25	5,3	72,2*	13,1	1,4	8,0	0,72	0,32	6,3	40,7*	11,9	12,8	28,3	0,41	0,20
9	6,9	3,2	5,8	2,6	81,5*	0,82	0,23	6,4	4,1	9,4	2,9	77,2*	0,77	0,30	6,3	4,9	20,1	6,3	62,6*	0,63	0,40
10	0,8	1,6	8,5	0,6	88,5*	0,89	0,33	1,4	2,6	16,9	1,4	77,7*	0,78	0,34	3,2	5,6	36,1	4,1	51,0*	0,51	0,39
11	11,1	3,8	80,6*	1,5	3,0	0,81	0,30	19,9	11,2	58,1*	7,4	3,3	0,58	0,26	26,9	16,0	46,8*	6,4	3,9	0,47	0,26
12	66,5*	5,9	3,4	18,9	5,3	0,67	0,25	74,3*	3,8	3,1	15,7	3,1	0,74	0,17	52,2*	7,9	4,4	18,4	17,1	0,52	0,19
13	69,3*	5,9	0,9	3,1	20,7	0,69	0,39	51,4*	20,8	0,8	4,3	22,8	0,51	0,39	71,1*	18,3	1,9	2,1	6,7	0,71	0,27
14	2,5	2,5	6,7	86,1*	2,3	0,86	0,26	1,9	4,9	9,9	81,2*	2,1	0,81	0,30	9,7	5,4	17,9	65,2*	1,9	0,65	0,33
15	2,9	1,7	91,5*	2,5	1,3	0,92	0,21	4,8	4,0	80,7*	6,9	3,6	0,81	0,31	5,2	3,0	85,5*	4,0	2,3	0,86	0,31
16	80,1*	1,1	14,4	3,0	1,4	0,80	0,26	80,9*	1,6	10,8	4,1	2,7	0,81	0,25	71,6*	2,5	15,2	6,8	3,9	0,72	0,30
17	4,6	2,3	3,1	85,5*	4,5	0,86	0,27	5,0	3,1	10,2	78,9*	2,7	0,79	0,27	7,8	5,7	12,2	67,0*	7,4	0,67	0,29
18	10,2	72,5*	5,8	9,5	2,0	0,73	0,43	12,2	60,3*	8,6	15,6	3,2	0,60	0,37	7,9	60,4*	4,5	21,2	6,0	0,60	0,32
19	0,9	2,4	6,5	0,4	89,8*	0,90	0,33	1,2	4,4	21,7	1,8	70,9*	0,71	0,40	1,5	8,1	37,6	3,3	49,4*	0,49	0,32
20	66,6*	1,4	0,7	21,2	10,1	0,67	0,40	61,5*	3,3	1,3	18,9	14,9	0,62	0,29	36,3*	3,4	2,8	30,2	27,3	0,36	0,29
21	24,9	0,7	0,8	72,7*	1,0	0,73	0,39	21,5	1,6	1,4	73,7*	1,8	0,74	0,27	29,7	1,9	4,1	59,9*	4,4	0,60	0,24
22	2,1	2,2	82,8*	10,8	2,1	0,83	0,39	6,1	6,8	60,4*	20,8	5,9	0.60	0,37	9,2	6,1	24,0*	34,8	25,9	0,24	0,37
23	0,6	0,5	2,3	89,0*	7,7	0,89	0,26	0,8	0,2	1,9	83,5*	13,6	0,84	0,35	1,4	1,0	2,3	80,5*	14,8	0,81	0,42
24	39,1	52,7*	3,5	4,2	0,6	0,53	0,33	39,2	44,6*	10,2	5,3	0,8	0,45	0,35	34,7	14,2*	26,1	24,2	0,8	0,14	0,32
25	2,1	2,1	88,6*	2,0	5,2	0,89	0,25	2,5	5,7	81,5*	2,8	7,5	0,82	0,27	7,2	8,9	64,7*	6,6	12,6	0,65	0,31
26	0,7	1,9	0,6	77,4*	19,4	0,77	0,31	0,6	6,7	1,6	54,2*	37,0	0,54	0,42	1,5	14,3	6,7	28,8*	48,8	0,29	0,38
27	73,0*	1,7	12,0	8,3	5,0	0,73	0,39	41,2*	3,9	19,7	26,3	9,0	0,41	0,35	27,4*	3,1	27,5	23,9	18,1	0,27	0,31
28	48,5*	37,7	2,0	10,1	1,7	0,49	0,31	36,4*	38,8	5,0	17,3	2,5	0,36	0,25	42,6*	23,5	5,6	26,0	2,4	0,43	0,34
29	40,6	2,0	7,4	0.3	49,7*	0,50	0,22	51,8	7,9	12,2	1,5	26,8*	0,27	0,19	41,4	3,9	28,6	2,6	23,5*	0,24	0,14
30	12,3	6,9	28,5*	43,8	8,5	0,29	0,32	7,8	4,9	15,4*	56,2	15,7	0,15	0,21	10,5	8,3	20,8*	49,4	11,0	0.21	0,31
	Mean					0,78	0,29						0,68	0,29						0,56	0,2

*Correct answer

TABLE A2
JAT 2: Reasoning

Item	White							Indian							Black						
	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}
	1	2	3	4	5			1	2	3	4	5			1	2	3	4	5		
1	0,3	0,6	0,6	0,7	97,9*	0,98	0,14	1,1	2,5	3,1	6,4	86,9*	0,87	0,30	5,2	5,6	27,6	10,4	51,2*	0,51	0,25
2	1.0	0,2	4,8	93,6*	0,3	0,94	0,19	0,8	0,0	1,6	97,3*	0,3	0,97	0,17	16,1	2,0	2,6	78,1*	1,2	0,78	0,31
3	90,5*	0,6	5,8	0,5	2,3	0,91	0,21	89,3*	0,4	5,5	0,4	4,5	0,89	0,29	41,3*	6,4	10,7	6,8	34,7	0,41	0,33
4	2,6	0,9	0,8	7,0	88,8*	0,89	0,21	3,9	1,9	3,0	11,5	79,7*	0,80	0,32	12,8	3,2	17,2	15,8	51,0*	0,51	0,38
5	5,1	83,7*	2,8	3,3	5,2	0,84	0,42	8,9	69,0*	8,3	4,9	8,9	0,69	0,39	10,7	31,3*	14,3	28,1	15,6	0,31	0,29
6	3,1	1,0	86,2*	1,5	8,1	0,86	0,28	4,7	2,4	62,9*	3,1	27,0	0,63	0,32	18,5	5,2	28,7*	20,5	27,1	0,29	0,20
7	79,3*	6,0	3,4	9,0	2,3	0,79	0,44	77,0*	7,1	3,8	9,3	2,8	0,77	0,42	46,6*	16,7	10,5	19,6	6,6	0,47	0,37
8	2,0	6,2	3,5	57,0*	31,3	0,57	0,42	15,4	15,0	1,5	46,2*	21,9	0,46	0,38	15,2	39,7	7,0	20,6*	17,5	0,21	0,19
9	2,3	5,0	76,6*	8,3	7,8	0,77	0,37	2,7	5,6	73,7*	9,3	8,8	0,74	0,36	3,7	16,9	59,7*	11,2	8,4	0,60	0,22
10	2,4	87,5*	4,1	1,1	4,9	0,88	0,27	4,7	71,3*	8,8	4,9	10,5	0,71	0,41	7,0	30,5*	14,5	18,9	29,2	0,31	0,32
11	12,6	5,4	4,0	71,0*	7,0	0,71	0,43	15,0	10,1	6,0	62,1*	6.8	0,62	0,47	20,6	21,4	15,8	33,1*	9,2	0,33	0,35
12	0,3	1,8	73,3*	18,6	6,1	0,73	0,37	1,3	0,8	57,9*	23,3	16,7	0,58	0,50	12,5	2,7	17,4*	19,3	48,1	0,17	0,30
13	11,5	4,3	0,6	7,8	75,9*	0,76	0,36	10,0	10,1	3,1	18,7	58,0*	0,58	0,30	36,1	5,2	3,8	32,9	22,2*	0,22	0,15
14	14,9	16,0	9,1	59,6*	0,5	0,60	0,43	10,4	31,1	17,6	39,6*	1,3	0,40	0,41	14,1	35,2	20,8	22,6*	7,3	0,23	0,25
15	70,8*	8,5	5,4	5,8	9,5	0,71	0,38	47,5*	13,2	14,6	16,8	7,8	0,48	0,43	34,0*	26,6	15,3	14,9	9,1	0,34	0,29
16	62,5*	18,7	4,0	9,0	5,7	0,63	0,36	43,1*	20,2	6,3	15,2	7,1	0,43	0,30	26,3*	29,2	10,0	25,9	8,6	0,26	0.24
17	0,9	16,0	0,9	66,8*	15,5	0,67	0,31	2,2	26,8	6,7	35,4*	28,9	0,35	0,37	8,3	51,8	10,5	11,7*	17,7	0,12	0,17
18	2,8	0,7	0,6	7,9	88,1*	0,88	0,22	20,8	1,2	1,1	12,3	64,5*	0,65	0,38	38,3	16,1	11,9	14,1	19,6*	0,20	0,18
19	7,7	14,2	23,5	7,6	47,0*	0,47	0,42	13,2	15,1	35,2	7,8	28,7*	0,29	0,43	9,8	14,7	40,1	19,3	16,2*	0,16	0,10
20	48,4*	1,2	28,6	4,5	17,4	0,48	0,35	26,9*	6,1	44,6	10,0	12,4	0,27	0,36	9,4*	19,4	55,7	8,4	7,1	0,09	0,15
21	7,5	20,3	38,8*	14,0	19,4	0,39	0,34	17,5	16,5	36,6*	10,2	19,2	0,37	0,31	12,6	26,0	20,5*	13,0	27,8	0,21	0,13
22	18,0	48,3*	9,7	10,2	13,8	0,48	0,42	14,9	43,3*	15,7	12,9	13,2	0,43	0.44	27,6	20,8*	18,0	22,6	11,0	0,21	0,23
23	32,4	3,9	47,1*	4,2	12,5	0,47	0,40	38,9	4,5	41,9*	5,0	9,7	0,42	0,43	52,0	5,9	25,1*	6,9	10,1	0,25	0,21
24	38,5*	15,8	16,5	11,6	17,5	0,39	0,46	21,6*	23,3	14,7	15,7	24,8	0,22	0,35	7,7*	14,7	10,1	13,7	53,8	0,08	0,12
25	1,2	42,8*	10,6	38,7	6,7	0,43	0,53	1,5	27,1*	19,1	46,1	6,2	0,27	0,47	4,7	7,8*	42,4	35,9	9,1	0,08	0,13
26	36,1	32,5*	1,2	14,8	15,4	0,33	0,48	37,3	25,3*	2,4	24,7	10,2	0,25	0,38	34,9	12,4*	11,5	27,2	14,1	0,12	0,11
27	17,3	19,6	12,5	43,7*	6,9	0,44	0,38	22,0	20,6	18,0	28,7*	10,7	0,29	0,21	56,6	11,6	11,9	14,8*	5,1	0,15	0,13
28	21,5	15,1	19,0	10,7	33,7*	0,34	0,51	28,6	11,2	25,1	15,7	19,3*	0,19	0,50	39,7	14,8	21,1	20,4	4,0*	0,04	0,06
29	31,1	38,6*	16,2	11,1	3,0	0,39	0,54	33,1	27,2*	25,2	10,2	4,3	0,27	0,45	24,3	12,7*	40,7	13,6	8,7	0,13	0,13
30	16,9	16,9	33,9*	20,1	12,3	0,34	0,44	13,5	15,4	28,9*	28,4	13,7	0,29	0,32	28,8	12,1	15,5*	30,5	13,0	0,16	0,22
Mean						0,63	0,37						0,50	0,37						0,26	0,21

*Correct answer

TABLE A3
JAT 3: Number Ability

Item	White							Indian							Black						
	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}
	1	2	3	4	5			1	2	3	4	5			1	2	3	4	5		
1	6,9	1,5	0,2	0,0	91,4*	0,91	0,20	8,5	2,2	0,8	0,1	88,5*	0,89	0,30	12,5	12,1	3,5	1,4	70,5*	0,71	0,30
2	2,2	3,8	1,9	85,2*	6,9	0,85	0,32	2,1	4,2	1,3	87,3*	5,1	0,87	0,28	3,4	13,9	2,0	67,7*	12,9	0,68	0,33
3	84,7*	1,3	2,7	1,8	9,5	0,85	0,25	84,8*	1,3	2,5	1,6	9,6	0,85	0,21	73,6*	3,2	5,2	3,8	14,2	0,74	0,32
4	0,9	83,8*	2,5	0,9	12,0	0,84	0,33	2,7	78,4*	2,4	1,0	15,4	0,78	0,35	7,9	55,9*	6,6	3,4	26,2	0,56	0,42
5	1,0	87,2*	4,4	2,4	5,0	0,87	0,31	1,2	86,7*	3,9	3,8	4,4	0,87	0,29	4,0	64,0*	8,5	9,2	14,3	0,64	0,39
6	5,6	3,8	76,7*	4,1	9,7	0,77	0,44	6,0	3,9	74,7*	4,5	10,9	0,75	0,39	10,1	9,8	54,0*	6,7	19,4	0,54	0,40
7	7,0	0,3	2,5	86,4*	3,8	0,86	0,21	6,5	0,5	2,7	86,2*	4,1	0,86	0,20	7,5	0,6	2,6	83,1*	6,2	0,83	0,25
8	82,5*	1,8	4,7	1,0	10,0	0,83	0,39	81,0*	1,6	6,0	0,8	10,6	0,81	0,41	63,3*	4,4	12,3	2,3	17,8	0,63	0,38
9	68,4*	2,2	7,1	4,9	17,4	0,68	0,36	69,6*	1,6	7,9	4,6	16,3	0,70	0,41	44,5*	7,9	15,9	6,4	25,3	0,45	0,44
10	5,0	1,3	6,3	73,4*	14,0	0,73	0,37	7,7	1,5	8,2	69,4*	13,1	0,69	0,42	7,6	5,8	14,4	49,3*	22,9	0,49	0,42
11	0,9	2,5	78,5*	14,7	3,5	0,79	0,40	1,5	2,8	74,2*	16,8	4,7	0,74	0,42	3,6	6,8	55,8*	24,3	9,5	0,56	0,35
12	13,6	7,9	8,8	16,7	53,1*	0,53	0,39	12,8	8,2	11,8	14,8	52,5*	0,53	0,38	17,2	15,0	11,9	17,8	38,2*	0,38	0,31
13	64,0*	4,4	7,3	11,2	12,9	0,64	0,39	64,9*	4,8	5,2	15,0	10,2	0,65	0,28	41,8*	8,6	13,0	17,0	19,7	0,42	0,39
14	0,6	88,7*	1,3	0,7	8,7	0,89	0,25	0,7	87,5*	1,4	0,9	9,4	0,88	0,23	3,2	71,5*	5,8	2,2	17,2	0,72	0,42
15	5,6	5,1	8,1	51,7*	29,4	0,52	0,46	7,6	5,6	5,8	55,9*	25,0	0,56	0,54	18,0	13,6	12,1	23,5*	32,8	0,24	0,25
16	20,4	9,6	7,6	12,5	49,9*	0,50	0,32	17,3	14,7	13,5	10,7	43,9*	0,44	0,40	12,5	22,7	22,4	15,7	26,8*	0,27	0,26
17	1,0	74,8*	3,5	1,7	19,0	0,75	0,40	0,8	74,3*	2,9	1,7	20,3	0,74	0,43	4,1	50,2*	14,3	5,5	26,0	0,50	0,48
18	3,8	6,1	9,3	63,3*	17,6	0,63	0,42	4,1	5,6	9,0	63,0*	18,2	0,63	0,46	11,6	12,8	15,0	33,9*	26,7	0,34	0,44
19	19,0	9,0	16,5	6,2	49,3*	0,49	0,45	15,5	6,7	19,5	9,9	48,4*	0,48	0,39	11,9	9,3	31,1	16,9	30,7*	0,31	0,33
20	11,9	49,7*	5,7	11,6	21,2	0,50	0,47	7,7	64,2*	7,2	8,2	12,6	0,64	0,48	16,0	30,0*	12,3	18,9	22,8	0,30	0,27
21	71,7*	4,9	5,8	5,6	12,0	0,72	0,45	67,1*	4,8	6,5	9,9	11,7	0,67	0,55	30,7*	14,1	19,9	17,5	17,8	0,31	0,40
22	10,0	8,3	57,4*	4,7	19,7	0,57	0,50	15,1	9,3	54,0*	6,0	15,6	0,54	0,40	20,6	14,3	27,7*	16,6	20,8	0,28	0,35
23	16,1	51,2*	8,9	5,7	18,0	0,51	0,45	14,8	57,2*	6,5	8,2	13,4	0,57	0,45	12,3	29,3*	17,9	20,3	20,2	0,29	0,31
24	6,6	9,7	42,9*	13,6	27,2	0,43	0,51	10,3	11,0	39,2*	14,3	25,3	0,39	0,50	16,4	24,5	22,1*	13,8	23,2	0,22	0,22
25	11,8	27,0	26,4*	10,4	24,4	0,26	0,39	10,9	25,4	26,0*	12,2	25,4	0,26	0,33	10,5	26,3	22,3*	20,4	20,4	0,22	0,13
26	24,9*	14,1	16,5	8,3	36,2	0,25	0,43	23,1*	18,3	16,7	10,2	31,6	0,23	0,46	10,6*	26,2	22,1	15,1	26,0	0,11	0,18
27	13,4	7,9	41,8*	5,6	31,3	0,42	0,44	15,0	7,7	45,8*	8,5	23,0	0,46	0,46	18,0	20,9	21,8*	13,1	26,3	0,22	0,15
28	3,8	11,3	17,7	49,9*	17,2	0,50	0,46	7,1	10,9	18,4	44,4*	19,1	0,44	0,43	14,3	20,8	26,4	21,0*	17,4	0,21	0,15
29	5,7	8,5	17,0	32,0*	36,8	0,32	0,42	8,6	8,9	22,7	33,9*	25,9	0,34	0,36	11,0	15,6	29,1	19,4*	24,8	0,19	0,17
30	16,4	30,6*	22,3	11,4	19,2	0,31	0,42	13,2	24,1*	27,5	16,4	18,8	0,24	0,38	20,0	17,2*	20,6	16,2	25,9	0,17	0,20
Mean						0,62	0,38						0,61	0,38						0,41	0,31

*Correct answer

TABLE A 4
JAT 4: Synonyms

Item	White							Indian							Black						
	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}
	1	2	3	4	5			1	2	3	4	5			1	2	3	4	5		
1	1,2	95,9*	1,6	0,6	0,7	0,96	0,26	3,4	79,4*	4,9	2,0	10,3	0,79	0,42	7,9	35,4*	22,7	19,5	14,6	0,35	0,29
2	2,0	1,4	0,8	2,3	93,5*	0,94	0,30	4,1	2,4	2,6	3,3	87,7*	0,88	0,40	20,7	19,3	15,9	13,4	30,7*	0,31	0,17
3	4,0	2,2	88,1*	2,0	3,8	0,88	0,32	10,9	4,0	67,7*	8,6	8,9	0,68	0,35	20,7	21,3	32,6*	10,3	15,0	0,33	0,30
4	8,9	0,9	1,0	88,0*	1,0	0,88	0,24	23,3	0,3	0,8	73,9*	1,7	0,74	0,30	27,1	3,3	8,8	57,8*	2,9	0,58	0,32
5	87,5*	4,2	1,2	3,3	3,8	0,88	0,33	77,2*	3,0	3,9	7,5	8,5	0,77	0,38	33,4*	41,9	5,7	7,5	11,5	0,33	0,36
6	3,3	3,7	2,7	1,9	88,4*	0,88	0,31	6,5	15,9	10,8	2,2	64,6*	0,65	0,27	5,0	16,4	29,4	7,2	41,9*	0,42	0,25
7	5,7	1,2	83,0*	8,6	1,5	0,83	0,26	19,2	6,9	19,1*	49,0	5,9	0,19	0,33	32,0	20,8	30,2*	9,9	7,1	0,30	0,32
8	83,4*	2,9	6,4	4,0	3,2	0,83	0,18	44,2*	13,5	31,5	5,6	5,2	0,44	0,12	19,2*	28,3	14,7	17,0	20,8	0,19	0,09
9	3,4	12,0	1,0	73,3*	10,2	0,73	0,36	4,5	24,7	4,5	49,6*	16,7	0,50	0,35	13,7	9,5	8,3	21,4*	47,0	0,21	0,23
10	8,6	65,1*	11,1	2,0	13,2	0,65	0,38	12,8	63,3*	8,5	5,5	10,0	0,63	0,49	17,7	20,5*	24,1	16,5	21,3	0,21	0,15
11	7,5	78,7*	1,4	3,0	9,3	0,79	0,43	5,1	58,5*	2,8	4,6	28,9	0,59	0,45	11,6	12,3*	20,1	18,2	37,8	0,12	0,20
12	4,6	8,0	4,8	63,3*	19,4	0,63	0,28	6,7	13,2	10,4	39,9*	29,8	0,40	0,29	18,9	9,6	14,4	45,4*	11,7	0,45	0,27
13	11,7	1,5	75,3*	2,1	9,5	0,75	0,24	3,2	1,5	48,1*	15,5	31,7	0,48	0,24	8,8	6,3	60,2*	12,9	11,6	0,60	0,22
14	68,3*	4,4	10,2	8,4	8,6	0,68	0,41	71,4*	4,1	10,2	4,6	9,7	0,71	0,39	8,1*	35,0	19,5	20,0	17,4	0,08	0,17
15	7,7	4,8	3,4	10,8	73,3*	0,73	0,44	3,8	8,3	3,8	29,4	54,6*	0,55	0,50	12,2	7,1	12,4	53,8	14,5*	0,15	0,23
16	8,8	7,3	69,1*	11,2	3,6	0,69	0,44	10,1	11,3	57,2*	11,9	9,5	0,57	0,47	12,3	23,3	18,2*	21,5	24,7	0,18	0,16
17	10,6	7,5	7,9	65,3*	8,8	0,65	0,51	7,7	12,4	16,3	38,1*	25,5	0,38	0,50	17,0	25,1	21,1	13,9*	23,0	0,14	0,15
18	14,9	5,5	6,9	3,3	69,3*	0,69	0,37	26,9	11,7	10,6	8,2	42,6*	0,43	0,43	20,0	14,9	25,0	15,1	25,1*	0,25	0,22
19	38,9*	2,9	16,8	2,2	39,1	0,39	0,34	62,0*	3,6	5,4	5,1	23,9	0,62	0,37	16,2*	18,2	25,9	21,3	18,5	0,16	0,11
20	27,8	3,6	4,6	59,0*	5,0	0,59	0,38	56,5	5,3	2,6	26,3*	9,4	0,26	0,44	46,2	12,0	17,0	16,2*	8,6	0,16	0,21
21	6,9	54,7*	4,9	13,2	20,3	0,55	0,42	5,7	56,7*	12,3	1,8	23,4	0,57	0,44	7,6	18,0*	20,7	21,2	32,6	0,18	0,10
22	12,8	7,7	29,7	4,9	44,9*	0,45	0,33	22,5	14,4	32,3	4,0	26,8*	0,27	0,35	23,0	14,4	23,7	20,4	18,5*	0,19	0,21
23	5,7	51,9*	14,2	2,5	25,8	0,52	0,54	9,7	67,6*	3,7	4,2	14,7	0,68	0,44	13,3	17,8*	29,9	26,3	12,7	0,18	0,23
24	50,9*	7,4	22,1	4,0	15,6	0,51	0,58	17,3*	2,2	51,1	1,2	28,2	0,17	0,48	9,6*	14,9	31,7	16,5	27,3	0,10	0,11
25	61,7*	8,1	15,1	4,6	10,6	0,62	0,36	41,9*	9,5	19,1	11,7	17,8	0,42	0,50	22,5*	15,3	19,5	19,0	23,8	0,23	0,26
26	9,1	7,6	54,0*	13,0	16,3	0,54	0,31	19,6	14,1	61,5*	2,6	2,3	0,62	0,47	15,0	29,0	29,5*	11,1	15,4	0,30	0,31
27	14,5	25,8	40,3*	11,6	7,8	0,40	0,32	39,1	15,4	26,8*	10,2	8,5	0,27	0,33	8,6	13,4	20,6*	44,0	13,4	0,21	0,19
28	7,5	12,7	12,9	26,8	40,1*	0,40	0,47	7,3	11,1	26,0	12,5	43,1*	0,43	0,43	18,3	15,0	26,0	15,8	24,9*	0,25	0,26
29	23,5	45,1*	7,2	14,5	9,7	0,45	0,41	17,7	43,7*	7,7	18,7	12,2	0,44	0,39	20,7	20,9*	13,8	22,7	21,9	0,21	0,20
30	7,4	6,0	12,3	63,3*	11,0	0,63	0,52	9,2	5,6	11,4	67,1*	6,8	0,67	0,47	17,8	27,4	26,0	20,0*	8,7	0,20	0,19
	Mean					0,67	0,36						0,52	0,39						0,25	0,21

*Correct answer

TABLE A5
JAT 6: Spatial 2 D

Item	White							Indian							Black						
	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}
	1	2	3	4	5			1	2	3	4	5			1	2	3	4	5		
1	92,6*	0,3	5,5	0,3	1,3	0,93	0,26	86,2*	2,5	7,0	0,8	3,5	0,86	0,27	79,4*	3,9	7,9	2,7	6,2	0,79	0,27
2	2,3	1,4	1,8	7,0	87,5*	0,88	0,34	4,7	4,0	4,7	12,6	73,9*	0,74	0,35	4,6	7,6	5,5	17,8	64,5*	0,65	0,36
3	90,5*	0,4	0,2	8,0	0,9	0,91	0,25	71,9*	5,9	0,4	15,6	6,2	0,72	0,47	57,7*	8,0	1,9	21,4	11,0	0,58	0,43
4	1,5	92,5*	0,4	4,8	0,9	0,93	0,35	5,4	73,7*	2,5	13,9	4,5	0,74	0,51	5,4	62,3*	5,3	16,7	10,3	0,62	0,46
5	2,8	82,2*	1,4	9,9	3,7	0,82	0,30	7,3	68,0*	3,3	13,6	7,8	0,68	0,30	5,3	56,2*	5,9	19,6	13,0	0,56	0,33
6	0,9	4,0	11,0	83,6*	0,5	0,84	0,33	2,8	9,8	14,8	70,8*	1,9	0,71	0,24	3,9	11,0	16,4	63,1*	5,6	0,63	0,26
7	3,7	3,9	88,1*	1,4	2,8	0,88	0,38	13,0	8,7	65,6*	6,5	6,3	0,66	0,53	18,5	8,5	48,1*	13,5	11,4	0,48	0,44
8	4,3	0,6	2,0	93,0*	0,2	0,93	0,32	15,6	2,6	3,6	76,5*	1,7	0,77	0,43	24,8	3,0	4,0	66,0*	2,2	0,66	0,44
9	4,8	1,0	12,3	81,8*	0,2	0,82	0,43	6,1	4,1	26,7	62,6*	0,5	0,63	0,37	6,5	5,3	35,5	49,5*	3,1	0,50	0,36
10	2,2	85,5*	2,9	8,0	1,4	0,86	0,34	11,5	66,2*	5,6	12,0	4,7	0,66	0,52	12,9	53,7*	10,1	13,0	10,4	0,54	0,49
11	5,2	77,2*	3,7	1,7	12,2	0,77	0,37	5,5	63,2*	5,8	6,9	18,7	0,63	0,32	7,4	49,1*	10,9	6,9	25,8	0,49	0,27
12	1,2	15,7	2,3	0,8	80,0*	0,80	0,42	3,4	24,9	4,7	2,2	64,9*	0,65	0,26	6,4	21,6	8,6	5,5	58,0*	0,58	0,30
13	7,3	4,2	80,7*	3,9	3,8	0,81	0,52	21,0	7,7	53,4*	5,5	12,5	0,53	0,59	24,5	11,8	38,4*	9,5	15,9	0,38	0,54
14	1,9	12,8	2,9	0,7	81,7*	0,82	0,41	2,3	20,6	6,1	9,2	61,9*	0,62	0,47	2,8	28,4	6,9	11,3	50,6*	0,51	0,45
15	9,9	14,9	72,4*	1,1	1,6	0,72	0,42	14,5	22,8	45,1*	10,0	7,6	0,45	0,41	16,0	19,4	38,7*	17,3	8,8	0,39	0,34
16	2,9	5,1	15,4	3,3	73,3*	0,73	0,27	4,7	6,4	19,2	6,6	63,1*	0,63	0,33	5,5	9,7	19,6	10,0	55,2*	0,55	0,39
17	2,5	2,9	8,6	4,2	81,8*	0,82	0,45	5,5	5,6	16,9	13,8	58,2*	0,58	0,49	6,2	7,6	25,6	18,1	42,5*	0,43	0,46
18	72,2*	12,8	1,8	10,0	3,2	0,72	0,39	56,0*	11,1	6,1	20,5	6,3	0,56	0,27	44,8*	14,8	8,9	19,1	12,3	0,45	0,32
19	50,8*	6,8	18,9	0,9	22,6	0,51	0,39	33,8*	8,2	19,9	8,3	29,8	0,34	0,31	24,2*	7,1	21,5	11,6	35,6	0,24	0,28
20	2,6	88,7*	5,1	2,2	1,4	0,89	0,30	12,2	64,8*	9,7	5,4	7,9	0,65	0,49	11,4	48,7*	11,3	14,0	14,7	0,49	0,47
21	23,6	1,8	3,9	66,2*	4,5	0,66	0,39	29,1	5,4	12,3	43,4*	9,8	0,43	0,46	28,2	7,9	15,5	37,0*	11,5	0,37	0,29
22	85,8*	1,5	8,3	3,6	0,8	0,86	0,42	65,4*	4,7	14,3	12,5	3,1	0,65	0,52	55,1*	8,0	19,0	13,9	4,0	0,55	0,47
23	1,5	15,4	79,1*	3,1	0,9	0,79	0,47	4,4	29,1	54,8*	7,2	4,4	0,55	0,45	7,3	28,6	46,7*	10,9	6,5	0,47	0,41
24	4,7	7,4	66,3*	10,1	11,6	0,66	0,53	14,5	19,5	37,4*	10,0	18,6	0,37	0,54	15,7	26,1	27,2*	8,0	23,0	0,27	0,49
25	3,6	13,3	3,7	4,7	74,6*	0,75	0,53	11,2	20,3	9,6	13,7	45,2*	0,45	0,52	10,5	24,0	12,9	19,7	32,9*	0,33	0,40
26	20,1	3,1	1,1	1,9	73,7*	0,74	0,34	27,7	4,2	4,1	10,6	53,3*	0,53	0,45	33,7	4,8	8,0	15,6	38,0*	0,38	0,38
27	27,4	11,1	60,3*	0,7	0,6	0,60	0,48	36,7	16,5	42,9*	1,4	2,5	0,43	0,42	42,8	24,3	25,0*	4,5	3,4	0,25	0,22
28	9,0	81,4*	3,0	5,4	1,3	0,81	0,40	14,0	56,1*	10,9	15,9	3,0	0,56	0,51	17,0	45,7*	12,6	19,2	5,4	0,46	0,43
29	77,0*	11,1	1,1	9,9	0,9	0,77	0,47	49,4*	27,3	2,9	17,0	3,5	0,45	0,53	34,4*	25,7	6,5	27,3	6,1	0,34	0,48
30	6,5	22,1	7,4	63,2*	0,8	0,63	0,50	9,1	32,2	14,3	42,2*	2,2	0,42	0,27	11,4	32,7	18,0	34,6*	3,3	0,35	0,24
31	51,6*	41,4	0,9	5,0	1,1	0,52	0,38	34,2*	45,8	7,8	9,6	2,6	0,34	0,25	35,8*	37,9	8,9	13,1	4,3	0,36	0,17
32	5,9	4,7	76,4*	5,8	7,2	0,76	0,41	10,0	20,0	48,6*	12,1	9,3	0,49	0,41	10,3	23,5	39,1*	16,7	10,4	0,39	0,35
33	4,0	9,1	10,6	68,8*	7,5	0,69	0,58	7,6	12,1	31,6	34,7*	14,0	0,35	0,53	10,6	16,2	35,5	25,1*	12,6	0,25	0,45
34	7,1	58,5*	20,0	7,5	6,9	0,59	0,37	21,7	38,7*	19,8	12,8	7,0	0,39	0,38	21,0	31,8*	20,9	15,5	10,9	0,32	0,32
35	4,7	25,9	63,4*	5,7	0,2	0,63	0,52	10,8	37,8	37,4*	9,5	4,4	0,37	0,51	14,4	43,1	21,1*	11,8	9,6	0,21	0,36
36	9,7	61,8*	9,2	1,2	18,1	0,62	0,51	28,0	30,3*	15,6	3,9	22,2	0,30	0,41	22,5	25,1*	19,3	5,5	27,6	0,25	0,29
37	3,9	3,5	9,4	13,4	69,8*	0,70	0,53	16,6	9,2	15,7	20,7	37,8*	0,38	0,44	19,5	11,4	18,4	19,8	30,8*	0,31	0,45
38	3,4	25,6	4,8	61,4*	4,7	0,61	0,38	7,7	30,4	11,6	40,5*	9,9	0,41	0,31	8,6	30,1	17,3	31,1*	12,8	0,31	0,28
39	11,5	4,3	18,9	63,0*	2,2	0,63	0,39	32,4	7,3	19,2	37,6*	3,5	0,38	0,37	31,1	12,0	18,9	33,5*	4,4	0,34	0,29
40	39,3*	35,5	18,1	1,5	5,6	0,39	0,34	18,8*	41,3	22,3	5,6	12,0	0,19	0,32	12,4*	40,7	31,1	9,2	6,6	0,12	0,11
Mean						0,74	0,40						0,53	0,41						0,42	0,36

*Correct answer

TABLE A6:
JAT 7: Spatial 3-D

Item	White							Indian							Black						
	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}	%testees selecting each option					p	r_{it}
	1	2	3	4	5			1	2	3	4	5			1	2	3	4	5		
1	1,1	1,3	96,4*	0,6	0,6	0,96	0,18	4,0	3,4	87,9*	2,1	2,6	0,88	0,31	6,1	7,7	79,2*	1,8	5,2	0,79	0,34
2	0,3	2,0	10,1	1,4	86,2*	0,86	0,34	0,9	7,5	16,9	4,9	69,7*	0,70	0,43	5,5	11,6	21,2	6,6	55,1*	0,55	0,42
3	0,5	0,9	14,3	0,8	83,5*	0,84	0,40	3,0	3,5	22,9	2,6	68,1*	0,68	0,51	9,5	8,9	25,2	5,7	50,7*	0,51	0,44
4	6,7	86,3*	2,1	3,1	1,8	0,86	0,36	10,7	71,7*	3,2	9,8	4,6	0,72	0,47	14,1	60,7*	6,1	12,7	6,4	0,61	0,44
5	2,3	0,5	13,6	0,9	82,7*	0,83	0,31	3,0	2,0	22,1	1,6	71,2*	0,71	0,27	7,1	3,8	21,0	3,1	65,1*	0,65	0,22
6	1,9	0,3	86,2*	11,2	0,4	0,86	0,38	9,1	1,0	56,4*	31,9	1,5	0,56	0,48	12,9	1,7	54,1*	28,2	3,2	0,54	0,38
7	0,5	0,3	0,9	93,6*	4,7	0,94	0,31	4,5	1,0	3,2	78,2*	13,1	0,78	0,45	6,0	6,2	3,7	69,6*	14,5	0,70	0,46
8	88,8*	6,3	1,3	0,8	2,7	0,89	0,44	63,6*	16,5	4,6	6,6	8,6	0,64	0,50	55,0*	16,7	8,8	7,8	11,7	0,55	0,40
9	13,8	78,5*	1,7	5,6	0,4	0,79	0,52	24,4	45,3*	10,7	16,7	3,0	0,45	0,57	29,4	27,0*	18,3	19,1	6,2	0,27	0,44
10	55,8*	4,9	6,3	26,2	6,7	0,56	0,28	41,4*	14,5	13,2	26,4	4,4	0,41	0,38	32,5*	16,1	14,7	31,2	5,3	0,33	0,34
11	2,1	2,7	2,1	66,4*	26,7	0,66	0,54	6,4	9,6	9,7	39,9*	34,4	0,40	0,51	5,6	12,0	12,4	19,4*	50,6	0,19	0,32
12	19,4	70,5*	9,0	0,7	0,5	0,71	0,41	20,3	52,9*	18,4	6,5	1,9	0,53	0,41	16,5	41,0*	26,6	12,5	3,4	0,41	0,36
13	7,9	7,6	6,5	69,1*	8,9	0,69	0,38	14,9	13,0	21,3	42,7*	8,1	0,43	0,46	22,7	14,4	25,9	28,7*	8,3	0,29	0,37
14	1,1	7,7	6,7	22,5	62,0*	0,62	0,37	4,7	6,9	30,0	20,6	37,8*	0,38	0,35	9,5	4,8	50,6	9,9	25,1*	0,25	0,34
15	14,5	56,1*	12,6	2,2	14,6	0,56	0,42	41,0	33,4*	10,7	2,0	12,9	0,33	0,45	56,3	16,2*	11,2	5,3	11,0	0,16	0,33
16	15,3	15,7	5,9	9,9	53,1*	0,53	0,38	9,6	22,3	10,5	18,5	39,1*	0,39	0,33	10,9	18,6	17,6	20,1	32,8*	0,33	0,16
17	13,4	3,6	58,3*	1,9	22,8	0,58	0,46	18,9	10,8	36,0*	7,5	26,8	0,36	0,45	25,8	17,0	20,9*	12,0	24,3	0,21	0,24
18	58,4*	1,3	9,6	15,5	15,2	0,58	0,40	37,8*	3,6	10,0	20,0	28,6	0,38	0,36	21,0*	5,4	14,0	21,3	38,3	0,21	0,29
19	2,1	5,6	8,7	62,9*	20,6	0,63	0,53	6,5	3,9	10,8	39,5*	39,3	0,40	0,56	12,3	3,0	16,7	15,1*	52,9	0,15	0,38
20	6,1	6,8	82,4*	2,8	1,9	0,82	0,53	9,0	19,5	56,9*	6,8	7,7	0,57	0,48	11,2	23,5	42,5*	12,9	9,9	0,43	0,46
21	6,4	5,0	82,9*	2,1	3,6	0,83	0,45	18,1	8,0	59,1*	4,0	10,8	0,59	0,49	21,5	7,9	48,0*	3,4	19,2	0,48	0,51
22	80,1*	5,2	4,9	2,7	7,0	0,80	0,49	56,0*	7,0	11,3	5,0	20,6	0,56	0,53	40,2*	10,7	13,6	8,5	27,0	0,40	0,42
23	6,7	5,6	6,0	7,6	74,1*	0,74	0,48	10,4	13,1	6,5	20,0	50,0*	0,50	0,52	12,8	19,8	11,9	22,4	33,0*	0,33	0,49
24	11,0	79,4*	4,5	3,0	2,2	0,79	0,52	14,4	48,9*	22,5	7,9	6,3	0,49	0,54	9,4	32,5*	34,5	10,3	13,2	0,33	0,47
25	75,0*	6,2	5,7	8,6	4,6	0,75	0,52	46,1*	11,3	11,7	19,3	11,6	0,46	0,52	27,1*	15,8	17,5	23,8	15,8	0,27	0,48
26	6,9	7,2	16,6	61,8*	7,6	0,62	0,40	10,0	11,0	27,7	38,4*	12,9	0,38	0,37	10,1	13,9	31,3	30,1*	14,6	0,30	0,28
27	9,1	7,8	76,2*	5,7	1,1	0,76	0,56	20,1	24,8	44,2*	9,1	1,8	0,44	0,56	26,3	27,3	29,1*	14,4	2,8	0,29	0,55
28	9,9	5,7	10,8	70,4*	3,2	0,70	0,50	16,8	7,3	22,1	45,2*	8,7	0,45	0,49	20,7	10,7	25,2	33,8*	9,5	0,34	0,44
29	6,2	70,6*	8,0	10,6	4,6	0,71	0,53	11,3	46,1*	17,6	16,8	8,3	0,46	0,52	18,2	29,5*	20,6	24,3	7,5	0,30	0,39
30	8,3	7,6	7,7	1,4	75,0*	0,75	0,53	16,5	6,3	27,5	2,1	47,5*	0,48	0,55	22,3	9,4	34,2	4,2	30,0*	0,30	0,49
Mean						0,74	0,43						0,51	0,46						0,38	0,38

*Correct answer

TABLE A7:
JAT 10: Mechanical Insight

Item	White					p	r_{it}	Indian					p	r_{it}	Black					p	r_{it}
	%testees selecting each option							%testees selecting each option							%testees selecting each option						
	1	2	3	4	5			1	2	3	4	5			1	2	3	4	5		
1	0,9	95,9*	1,5	0,8	0,9	0,96	0,12	2,3	87,9*	2,1	2,6	5,1	0,88	0,20	18,1	57,2*	9,0	9,3	6,4	0,57	0,23
2	79,3*	0,6	14,4	2,1	3,6	0,79	0,36	50,8*	1,9	34,3	3,6	9,4	0,51	0,44	19,0*	8,1	52,7	7,7	12,5	0,19	0,23
3	5,3	3,7	84,5*	2,9	3,5	0,85	0,31	7,3	3,4	74,4*	5,4	9,5	0,74	0,31	10,0	11,4	53,3*	10,5	14,7	0,53	0,29
4	50,7	1,9	10,0	31,6*	5,9	0,32	0,27	53,2	2,6	10,0	23,3*	10,9	0,23	0,20	42,3	10,3	14,8	19,5*	13,1	0,20	0,15
5	3,4	82,1*	2,8	2,2	9,5	0,82	0,27	6,0	65,2*	3,3	3,5	22,0	0,65	0,32	16,6	41,8*	9,7	8,0	23,9	0,42	0,22
6	71,7*	11,2	3,7	3,6	9,8	0,72	0,39	51,0*	20,5	7,6	6,4	14,4	0,51	0,28	50,4*	16,5	11,0	9,5	12,6	0,50	0,18
7	1,7	0,6	82,4*	10,8	4,6	0,82	0,37	5,8	1,3	53,8*	23,2	16,0	0,54	0,38	8,0	4,5	37,2*	28,5	21,8	0,37	0,29
8	3,4	0,7	6,4	64,1	25,4*	0,25	0,19	8,3	2,4	17,6	45,3	26,6*	0,27	0,03	11,5	7,5	29,3	29,2	22,4*	0,22	0,01
9	3,9	0,9	13,8*	60,8	20,6	0,14	0,13	10,8	1,6	16,6*	45,5	25,4	0,17	0,08	13,4	7,7	17,6*	34,8	26,6	0,18	0,10
10	5,3	7,8	14,8	35,1*	37,2	0,35	0,32	11,9	20,4	13,1	23,2*	31,5	0,23	0,17	21,2	21,4	19,3	19,3*	18,8	0,19	0,13
11	53,9*	12,3	20,1	9,6	4,2	0,54	0,14	45,2*	12,6	28,6	8,4	5,2	0,45	0,14	40,1*	10,1	22,0	18,2	9,6	0,40	0,12
12	59,0*	14,3	12,6	7,7	6,5	0,59	0,42	25,6*	10,6	20,4	14,1	29,4	0,26	0,44	10,3*	11,4	20,2	17,5	40,4	0,10	0,17
13	1,0	0,9	90,4*	2,1	5,5	0,90	0,30	3,5	5,0	60,0*	5,6	25,9	0,60	0,42	7,4	9,7	28,0*	10,8	44,1	0,28	0,28
14	6,1	74,1*	3,0	3,3	13,5	0,74	0,35	10,6	51,9*	5,6	7,7	24,2	0,52	0,42	18,5	25,0*	15,7	13,5	27,4	0,25	0,27
15	55,9*	30,4	3,2	2,5	8,0	0,56	0,26	46,2*	26,9	6,3	10,2	10,4	0,46	0,30	36,1*	16,6	14,5	14,8	18,0	0,36	0,25
16	8,9	8,3	21,7	4,8	56,3*	0,56	0,25	14,3	10,4	22,9	3,6	48,9*	0,49	0,19	14,4	11,3	30,7	6,4	37,1*	0,37	0,24
17	9,9	6,2	49,0*	2,2	32,7	0,49	0,26	17,0	7,7	30,4*	3,0	42,0	0,30	0,32	19,6	11,3	26,8*	8,1	34,1	0,27	0,20
18	8,8	45,2*	16,7	19,5	9,8	0,45	0,33	12,7	31,2*	14,3	25,7	16,2	0,31	0,24	19,8	20,5*	20,2	22,3	17,1	0,21	0,22
19	18,6	18,8	17,1	37,9*	7,5	0,38	0,34	16,9	25,2	20,2	24,2*	13,4	0,2	0,20	18,8	22,8	20,3	19,9*	18,2	0,20	0,16
20	20,2	19,2	19,5	31,5*	9,6	0,32	0,43	23,1	20,3	20,3	20,0*	16,4	0,20	0,16	28,0	13,6	18,4	13,4*	26,5	0,13	0,05
21	9,0	34,1*	6,6	44,1	6,2	0,34	0,12	18,5	35,5*	14,5	17,1	14,4	0,36	0,17	17,3	30,3*	20,1	15,2	17,1	0,30	0,18
22	34,5	7,5	20,2	29,2	8,5*	0,09	0.07	30,8	14,2	20,2	21,7	13,1*	0,13	0,01	28,9	18,0	21,0	15,5	16,6*	0,17	0,17
23	7,7	29,9	15,3	17,1	29,9*	0,30	0,25	14,0	19,2	21,6	21,3	23,9*	0,24	0,22	17,3	15,9	25,0	23,5	18,3*	0,18	0,15
24	4,2	7,8	75,5*	7,4	5,1	0,76	0,36	10,5	17,4	48,1*	12,8	11,2	0,48	0,37	13,0	19,3	33,0*	17,2	17,5	0,33	0,26
25	2,1	3,1	62,9*	20,7	11,2	0,63	0,46	5,7	6,4	32,0*	38,6	17,3	0,32	0,36	8,9	13,0	27,8*	29,5	20,8	0,28	0,26
26	6,2	3,8	5,9	27,3	56,7*	0,57	0,28	14,2	8,6	14,4	19,4	43,4*	0,43	0,18	15,7	16,9	18,6	21,2	27,6*	0,28	0,12
27	7,3	6,5	5,5	58,0*	21,7	0,59	0,39	17,5	14,3	15,1	27,7*	25,4	0,28	0,33	14,4	18,5	19,7	17,7*	29,7	0,18	0,17
28	28,6	8,7	5,9	4,6	52,2*	0,52	0,21	32,8	9,9	9,9	10,2	37,1*	0,37	0,32	31,3	13,0	17,2	15,9	22,6*	0,23	0,27
29	8,5	31,9	32,6*	17,8	9,2	0,33	0,41	8,9	28,2	18,8*	21,1	23,0	0,19	0,22	5,7	28,4	19,5*	22,6	23,9	0,20	0,16
30	3,1	26,5	5,6	58,6*	6,2	0,59	0,27	5,5	24,8	9,0	53,4*	7,3	0,53	0,28	12,0	32,9	23,3	22,3*	9,6	0,22	0,19
31	9,0	37,2*	23,0	27,6	3,2	0,37	0,25	6,9	35,9*	30,2	23,6	3,4	0,36	0,18	9,6	27,7*	28,8	24,6	9,3	0,28	0,22
32	23,2	11,9	19,8	15,1	30,0*	0,30	0,30	19,8	12,7	26,2	12,9	28,4*	0,28	0,14	12,7	18,5	31,9	16,5	20,4*	0,20	0,17
33	72,1*	7,0	7,9	8,3	4,6	0,72	0,41	37,7*	16,0	22,5	11,6	12,2	0,38	0,34	25,6*	17,7	28,4	15,4	12,9	0,26	0,25
34	78,3*	5,5	6,0	7,2	3,0	0,78	0,36	38,4*	11,2	29,3	12,8	8,3	0,38	0,42	15,8*	15,5	39,0	17,3	12,4	0,16	0,20
35	16,0	7,5	46,9*	20,3	9,3	0,47	0,19	22,3	10,2	42,0*	11,4	14,2	0,42	0,22	18,2	16,3	31,3*	17,5	16,7	0,31	0,27
36	4,1	67,3*	14,6	11,4	2,7	0,67	0,37	13,2	47,1*	17,7	16,6	5,4	0,47	0,27	16,9	30,5*	22.8	18,9	10,9	0,31	0,26
37	39,2	20,0	11,5	17,7*	11,7	0,18	0,18	35,9	18,2	18,5	12,7*	14,6	0,13	0,07	20,3	24,5	26,9	17,7*	10,6	0,18	0,16
38	22,1*	58,8	4,0	7,8	7,3	0,22	0,14	22,0*	41,9	9,3	9,3	17,6	0,22	0,08	17,5*	28,4	17,9	15,1	21,1	0,18	0,19
39	21,3	55,9*	12,6	8,1	2,1	0,56	0,36	20,7	36,9*	19,4	14,4	8,6	0,37	0,25	21,7	27,1*	24,6	17,1	9,5	0,27	0,25
40	16,8	17,3	17,1	43,3*	5,5	0,43	0,45	23,0	27,1	20,6	20,7*	8,6	0,21	0,20	16,0	27,5	21,9	19,2*	15,4	0,19	0,17
41	52,0*	14,7	12,2	11,8	9,4	0,52	0,39	20,3*	20,5	23,2	23,0	12,9	0,20	0,15	12,9*	20,6	25,4	26,1	15,0	0,13	0,16
42	22,1	11,0	12,2	14,3	40,5*	0,41	0,36	21,8	19,9	19,5	15,0	23,8*	0,24	0,13	13,1	23,2	20,5	22,7	20,5*	0,21	0,15
Mean						0,52	0,29						0,37	0,24						0,26	0,19

*Correct answer

APPENDIX B

Bias indices based on Scheuneman's chi-square (χ^2_3), the full chi-square (χ^2_4), and the item characteristic curves (ICCs - for JAT 1 and JAT 2 only) for seven tests of the JAT

TABLE B1
JAT 1: Classification

Item number	Bias indices					
	Indian			Black		
	χ^2_3	χ^2_4	ICC	χ^2_3	χ^2_4	ICC
1	1,19	9,42*	0,39	1,56	11,08*	0,59
2	0,59	12,90*	0,43	4,24	62,90**	0,47
3	3,69	26,23**	0,17	1,79	19,64**	0,22
4	0,52	8,52	0,24	1,85	19,19**	0,18
5	5,67	32,10**	0,35	18,58**	66,80**	0,57
6	1,23	5,95	0,37	3,85	14,68**	0,20
7	0,70	4,19	0,07	0,25	2,00*	0,15
8	11,71**	44,85**	0,53	84,92**	264,28**	1,03
9	2,39	7,65	0,07	4,79	11,85*	0,13
10	6,04	26,32**	0,30	24,93**	87,72**	0,49
11	33,00**	108,77**	0,42	31,35**	86,01**	0,62
12	5,05	16,69**	0,19	8,69*	20,09**	0,30
13	23,39**	54,88**	0,51	2,94	10,30*	0,19
14	1,76	8,74	0,17	8,79*	32,71**	0,36
15	8,52*	47,81**	0,54	1,08	7,50	0,16
16	0,74	3,37	0,09	1,49	6,41	0,06
17	2,14	10,33*	0,08	6,72	33,47**	0,34
18	9,03*	24,86**	0,28	4,02	12,32*	0,20
19	18,64**	87,73**	0,37	33,55**	115,36**	0,64
20	2,67	6,44	0,14	36,50**	62,54**	0,58
21	1,43	5,08	0,17	3,88	7,92	0,32
22	28,75**	88,51**	0,38	116,38**	251,78**	0,68
23	2,09	9,46*	0,44	1,85	4,39	0,15
24	13,41*	19,82**	0,59	107,74**	135,54**	0,83
25	2,70	16,77**	0,21	19,27**	61,49**	0,45
26	38,96**	98,08**	0,63	106,11**	186,95**	0,92
27	77,21**	179,46**	0,50	67,43**	130,24**	0,85
28	25,57**	38,66**	0,29	8,31*	9,47*	0,16
29	69,87**	107,13**	0,54	72,10**	91,34**	0,62
30	30,06**	36,84**	0,21	2,37	10,35*	0,12

**p<0,01
*p<0,05
ICC>0,2 is probably significant

TABLE B2
JAT 2: Reasoning

Item number	Bias indices					
	Indian			Black		
	χ^2_3	χ^2_4	ICC	χ^2_3	χ^2_4	ICC
1	6,82	78,68**	0,28	61,73**	407,01**	0,92
2	1,50	27,87**	0,43	1,78	14,20**	0,21
3	1,46	8,87	0,01	66,91**	211,61**	0,87
4	4,84	26,05**	0,44	35,47**	98,94**	0,72
5	7,10	35,61**	0,14	41,45**	111,06**	0,39
6	35,57**	143,64**	0,45	143,09**	401,18**	1,25
7	1,32	5,29	0,04	11,00*	23,57**	0,22
8	8,13*	15,17**	0,13	46,99**	83,66**	0,83
9	0,76	2,37	0,15	7,00	27,60**	0,58
10	13,65**	53,77**	0,39	89,07**	220,77**	0,72
11	4,53	8,81	0,22	25,74**	49,52**	0,40
12	19,16**	34,47**	0,32	125,46**	232,58**	0,66
13	19,13**	56,26**	0,28	144,46**	300,22**	#
14	27,53**	54,92**	0,23	27,88**	51,52**	0,42
15	37,13**	88,17**	0,34	37,27**	83,79**	0,78
16	25,20**	56,16**	0,28	22,70**	41,41**	0,33
17	89,45**	172,55**	0,54	187,73**	286,21**	0,83
18	37,19**	136,04**	0,64	253,59**	559,56**	#
19	29,36**	45,91**	0,17	28,92**	41,74**	0,38
20	60,64**	90,95**	0,34	134,40**	171,62**	#
21	2,37	3,21	0,05	30,20**	39,12**	#
22	1,52	2,00	0,05	23,16**	29,66**	0,74
23	5,29	7,47	0,12	15,50**	17,16**	0,61
24	38,79**	47,44**	0,17	70,13**	75,44**	0,65
25	27,14**	34,12**	0,16	72,84**	78,24**	0,37
26	5,03	5,30	0,04	31,78**	37,81**	0,92
27	36,38**	41,23**	0,27	53,88**	55,90**	0,35
28	34,13**	35,80**	0,13	61,77**	71,49**	1,03
29	22,83**	25,76**	0,11	36,87**	51,66**	0,94
30	4,08	9,08	0,12	17,96**	32,38**	0,14

ICC>0,2 is probably significant
**p<0,01
*p<0,05
#Not possible to estimate item parameters

TABLE B3
JAT3: Number Ability

Item number	Bias indices			
	Indian		Black	
	χ^2_3	χ^2_4	χ^2_3	χ^2_4
1	0,41	2,40	4,74	22,94**
2	1,04	6,51	1,26	4,11
3	1,04	4,96	1,73	8,94
4	0,98	4,24	9,40*	28,80**
5	0,14	1,04	7,89*	26,50**
6	2,15	7,35	1,83	7,97
7	0,37	2,36	0,84	5,67
8	0,42	2,55	1,54	4,37
9	4,58	-13,75**	7,27	13,68**
10	0,31	0,73	1,68	2,97
11	0,42	1,23	5,73	16,00**
12	1,09	2,29	2,11	4,11
13	1,74	4,26	10,88*	18,96**
14	0,26	1,71	6,44	17,72**
15	-11,83**	-25,93**	18,51**	24,36**
16	0,93	1,44	14,45**	16,11**
17	0,79	3,10	15,11**	21,77**
18	1,69	3,61	20,24**	26,67**
19	1,20	2,67	7,56	-13,38**
20	-29,93**	-81,65**	12,27*	-17,70**
21	0,57	0,78	61,19**	78,56**
22	0,32	0,49	47,43**	53,36**
23	6,40	-16,79**	28,28**	-37,08**
24	1,73	3,18	17,99**	-32,67**
25	-10,03*	-12,60*	14,48**	-44,90**
26	2,84	4,35	19,06**	-39,21**
27	7,76	-15,93*	29,16**	-44,26**
28	3,99	7,15	54,73**	-68,32**
29	2,34	5,55	15,74**	-39,84**
30	3,71	4,49	7,60	-35,31**

**p<0,01 -Bias in favour of Indian pupils
*p<0,05

-Bias in favour of black pupils

TABLE B4
JAT 4: Synonyms

Item number	Bias indices			
	Indian		Black	
	χ^2_3	χ^2_4	χ^2_3	χ^2_4
1	14,63**	101,56**	132,00**	538,94**
2	1,58	17,10**	157,27**	555,15
3	21,72**	98,49**	95,69**	256,20
4	13,38**	52,10**	24,05**	93,23
5	5,96	25,64**	97,23**	226,92
6	36,07**	151,64**	82,00**	247,03
7	415,39**	824,90**	99,71**	253,04
8	122,80**	343,88**	232,04**	500,83
9	37,69**	98,13**	108,14**	221,41
10	0,45	0,98	91,71**	174,83
11	23,74**	73,47**	247,27**	437,74
12	48,19**	98,05**	8,46*	17,95
13	58,59**	156,25**	8,01*	28,86
14	3,86	12,29*	229,46**	369,86
15	16,43**	46,58**	129,96**	260,18
16	10,00*	21,20**	137,39**	240,40
17	58,53**	121,99**	137,69**	243,17
18	61,48**	129,48**	92,92**	172,92
19	62,89**	126,22**	56,74**	73,89
20	121,49**	197,39**	84,35**	126,79
21	2,27	5,48	79,51**	125,38
22	39,05**	58,00**	40,21**	60,27
23	34,03**	83,56**	47,87**	73,02
24	140,05**	211,89**	87,32**	130,68
25	36,52**	64,66**	69,95**	113,49
26	13,73**	35,67**	12,28**	18,54
27	24,00**	34,00**	27,80**	31,39
28	7,83*	13,06*	9,75*	16,23
29	0,56	1,27	31,49**	34,66
30	5,66	20,11**	88,51**	123,66

**p<0,01
*p<0,05

p<0,01 applies to all the values in this column

TABLE B5
JAT 6: Spatial 2-D

Item number	Bias indices			
	Indian		Black	
	χ^2_3	χ^2_4	χ^2_3	χ^2_4
1	1,38	9,98*	1,66	19,63**
2	0,93	2,64	1,79	6,35
3	4,06	10,72*	6,22	19,06**
4	3,10	5,87	2,62	5,38
5	0,51	1,31	2,72	6,21
6	1,29	6,28	0,60	1,66
7	0,92	2,24	4,20	16,22**
8	0,85	4,24	1,26	2,98
9	4,13	7,75	2,83	6,42
10	1,48	5,55	3,11	11,16*
11	3,18	6,44	2,99	10,64*
12	5,98	19,45**	5,65	13,92**
13	1,78	2,84	1,17	1,47
14	3,25	11,42*	1,47	5,79
15	5,39	11,11*	3,68	10,83*
16	2,36	6,86	6,48	19,90**
17	0,76	3,44	2,86	6,89
18	1,55	3,85	0,28	0,55
19	11,23*	17,80**	6,23	9,85*
20	6,95	15,25**	12,78**	26,59**
21	0,73	2,20	2,32	6,32
22	1,13	4,18	0,80	3,67
23	3,38	19,97**	1,30	3,23
24	1,74	2,20	1,70	2,82
25	2,37	8,75	2,37	9,54*
26	2,04	3,84	7,45	12,79*
27	4,91	10,28*	7,14	23,78**
28	5,22	10,18*	4,62	6,00
29	2,24	5,77	5,24	8,37
30	6,66	19,96**	2,56	6,16
31	4,19	7,34	1,18	5,92
32	11,89**	19,24**	9,26*	15,17**
33	9,74*	16,52**	4,10	10,75*
34	2,48	4,66	2,37	4,00
35	2,20	8,78	10,45*	18,48**
36	16,27**	22,66**	9,59*	17,44**
37	18,27**	22,48**	3,60	6,04
38	10,46*	11,14*	10,21*	11,86*
39	13,55**	15,08**	5,42	9,37
40	23,65**	24,81**	31,62**	37,23**

**p<0,01
*p<0,05

TABLE B6
JAT 7: Spatial 3-D

Item number	Bias indices			
	Indian		Black	
	χ^2_3	χ^2_4	χ^2_3	χ^2_4
1	0,60	6,25	1,32	7,82
2	1,78	3,45	3,47	6,72
3	1,41	5,74	1,49	13,01*
4	0,68	4,96	1,13	4,72
5	0,62	2,37	1,10	5,13
6	13,16**	35,50**	1,35	3,33
7	1,99	6,74	2,22	7,74
8	2,57	16,77**	0,47	8,90
9	7,04	18,56**	8,97*	23,70**
10	4,71	7,24	9,04*	14,00**
11	5,37	10,82*	15,12**	29,95**
12	1,91	5,34	2,43	6,30
13	5,37	12,12*	5,23	12,58*
14	6,41	16,98**	7,47	12,79*
15	1,32	2,53	10,81*	14,86**
16	1,27	2,64	2,88	5,17
17	0,69	0,94	7,95*	20,90**
18	3,17	4,72	8,66*	20,70**
19	3,81	6,28	13,28**	23,79**
20	4,21	16,18**	2,17	6,11
21	0,96	5,32	1,42	4,55
22	1,33	4,66	0,78	5,07
23	3,80	5,79	6,90	8,68
24	3,88	19,17**	5,78	8,56
25	1,71	5,32	4,50	6,36
26	4,32	9,28	4,90	7,50
27	2,74	13,84**	2,72	3,93
28	3,07	6,04	3,46	10,69*
29	2,78	4,42	3,56	7,10
30	2,06	8,73	3,76	9,79*

**p<0,01
*p<0,05

TABLE B7
JAT 10: Mechanical Insight

Item number	Bias indices			
	Indian		Black	
	χ^2_3	χ^2_4	χ^2_3	χ^2_4
1	2,73	31,50**	38,40**	264,24**
2	70,25**	155,18**	178,92**	345,28**
3	6,36	21,87**	20,96**	65,79**
4	16,02**	22,28**	9,57*	13,90**
5	16,47**	54,40**	34,45**	95,95**
6	29,97**	75,07**	8,77*	26,43**
7	59,01**	154,59**	50,32**	120,13**
8	4,51	5,76	4,73	6,13
9	4,59	5,65	2,16	5,52
10	21,74**	30,55**	17,14**	23,64**
11	9,17*	17,78**	6,72	12,87*
12	117,83**	191,63**	118,37**	186,67**
13	50,97**	200,85**	102,31**	334,65**
14	40,09**	85,32**	97,37**	195,22**
15	7,98*	15,67**	9,82*	16,47**
16	3,82	8,05	9,02*	14,89**
17	39,14**	60,06**	18,84**	31,58**
18	20,68**	33,01**	28,47**	40,51**
19	28,11**	41,44**	19,95**	27,04**
20	29,65**	40,61**	17,29**	23,75**
21	4,32	6,73	0,68	1,00
22	-15,48**	-17,99**	-16,00**	-19,57**
23	7,21	8,88	9,64*	10,92*
24	56,10**	125,80**	50,05**	79,81**
25	86,52**	155,51**	38,13**	64,36**
26	13,21**	22,66**	26,45**	37,91**
27	105,39**	176,55**	80,02**	109,60**
28	32,95**	46,86**	51,56**	61,01**
29	45,17**	52,04**	20,37**	26,46**
30	6,73	8,85	89,47**	107,68**
31	3,65	6,25	15,89**	19,02**
32	7,29	-11,34*	10,23*	-19,25**
33	139,54**	209,35**	148,53**	198,58**
34	172,00**	265,25**	243,73**	340,28**
35	18,74**	25,43**	40,87**	43,08**
36	73,79**	78,96**	79,84**	88,47**
37	21,47**	-40,40**	4,18	-26,61**
38	11,15*	-44,21**	20,55**	-32,68**
39	91,76**	93,34**	68,89**	70,50**
40	111,86**	-114,01**	47,62**	-56,85**
41	159,96**	161,93**	99,79**	105,97**
42	82,07**	-92,58**	41,90**	-57,04**

**p<0,01 -Bias in favour of Indian pupils
*p<0,05

-Bias in favour of black pupils

APPENDIX C

Item characteristic curves (ICCs) for JAT 1 and JAT 2

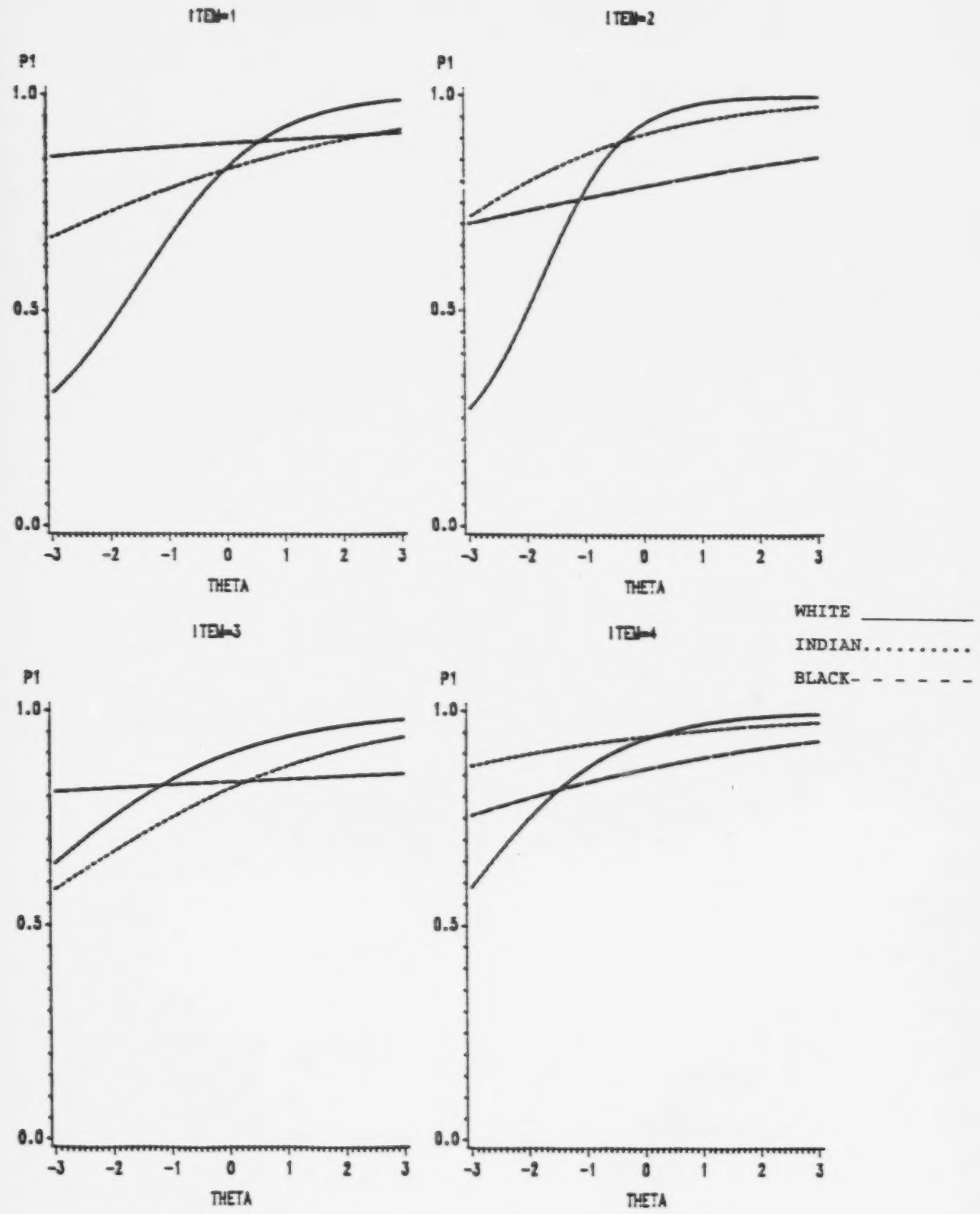


FIGURE C 1: JAT 1: ITEM CHARACTERISTIC CURVES OF THE ITEMS FOR WHITE, INDIAN AND BLACK PUPILS

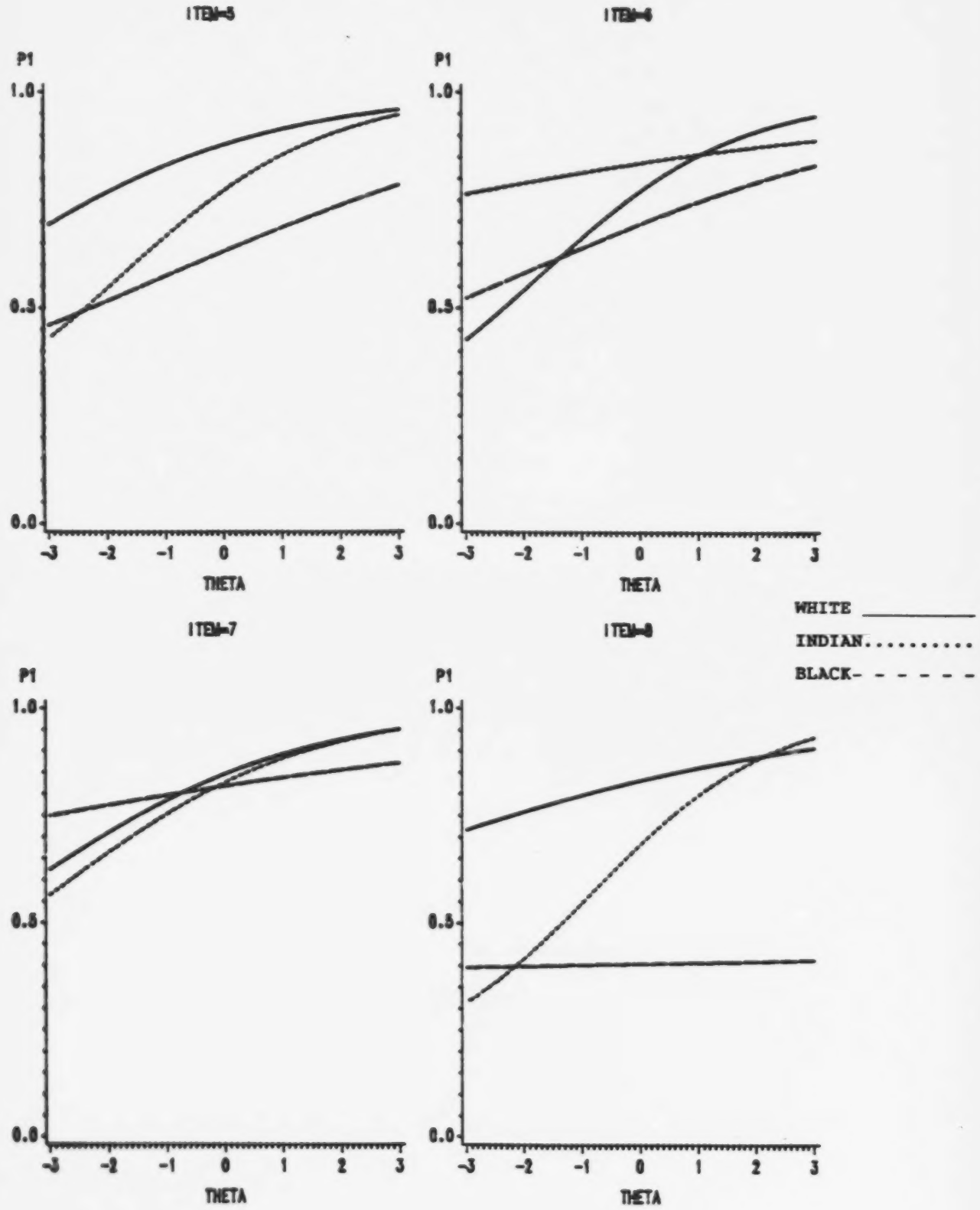


FIGURE C1: JAT 1: (CONTINUED)

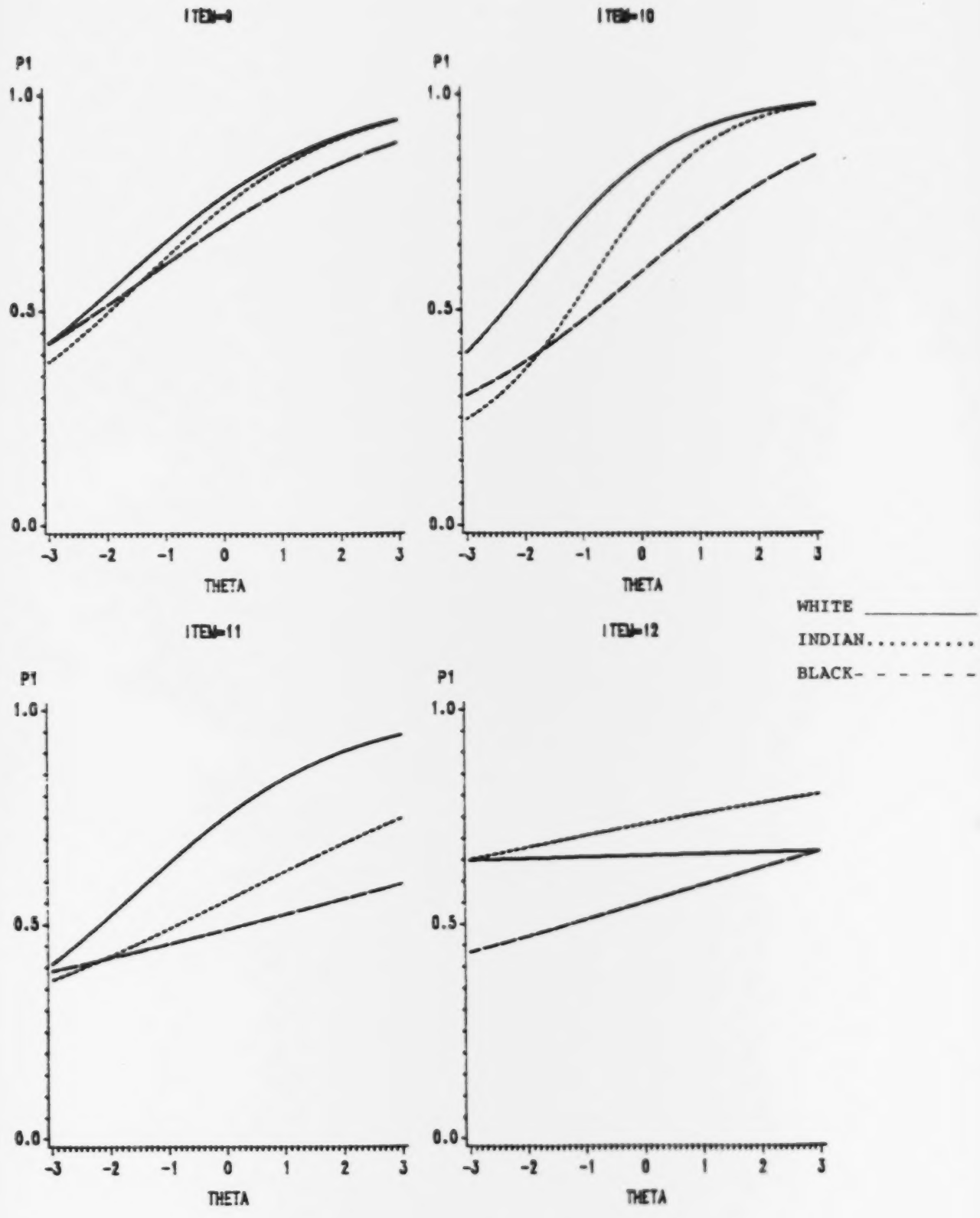


FIGURE C1: JAT 1: (CONTINUED)

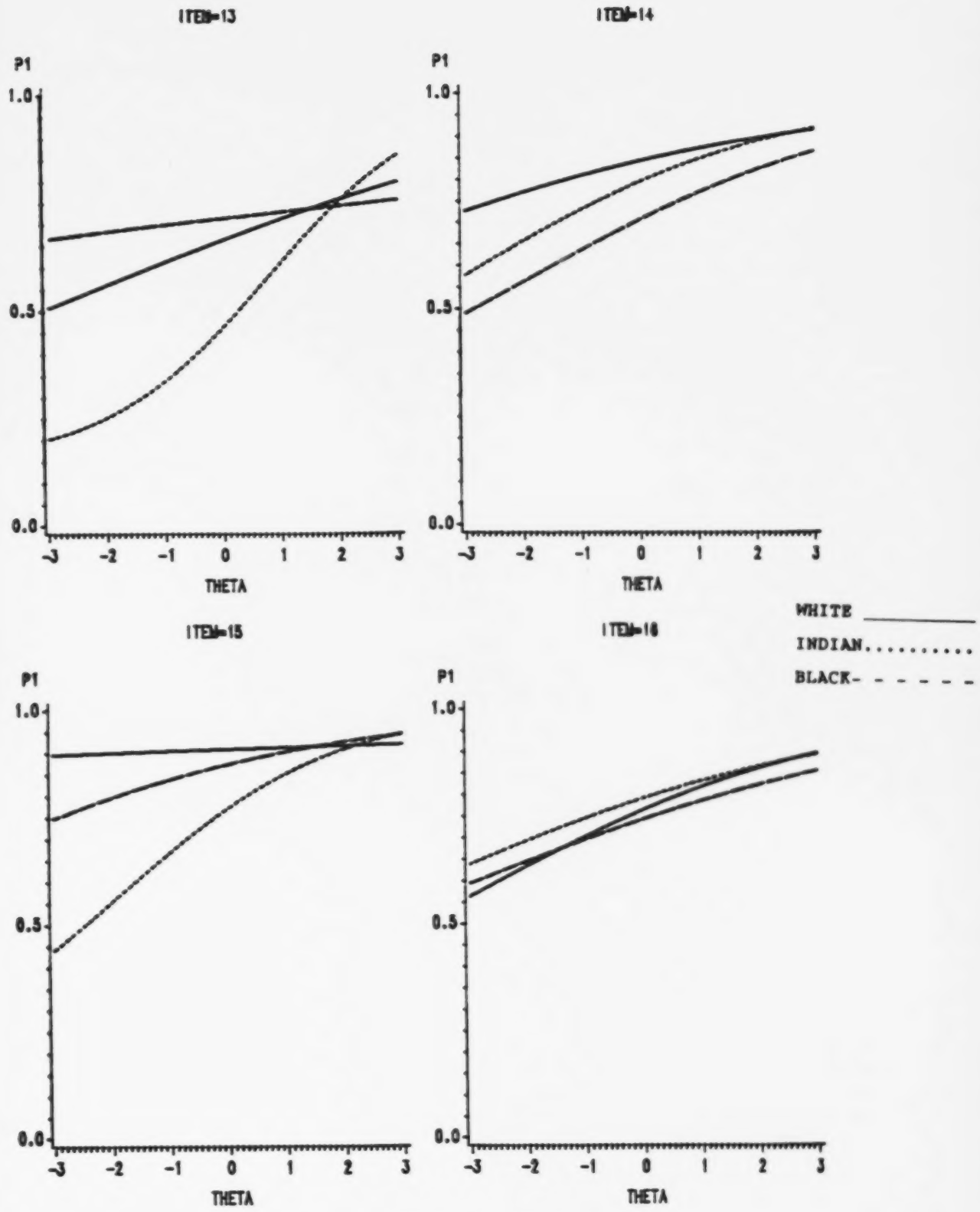


FIGURE C1: JAT 1: (CONTINUED)

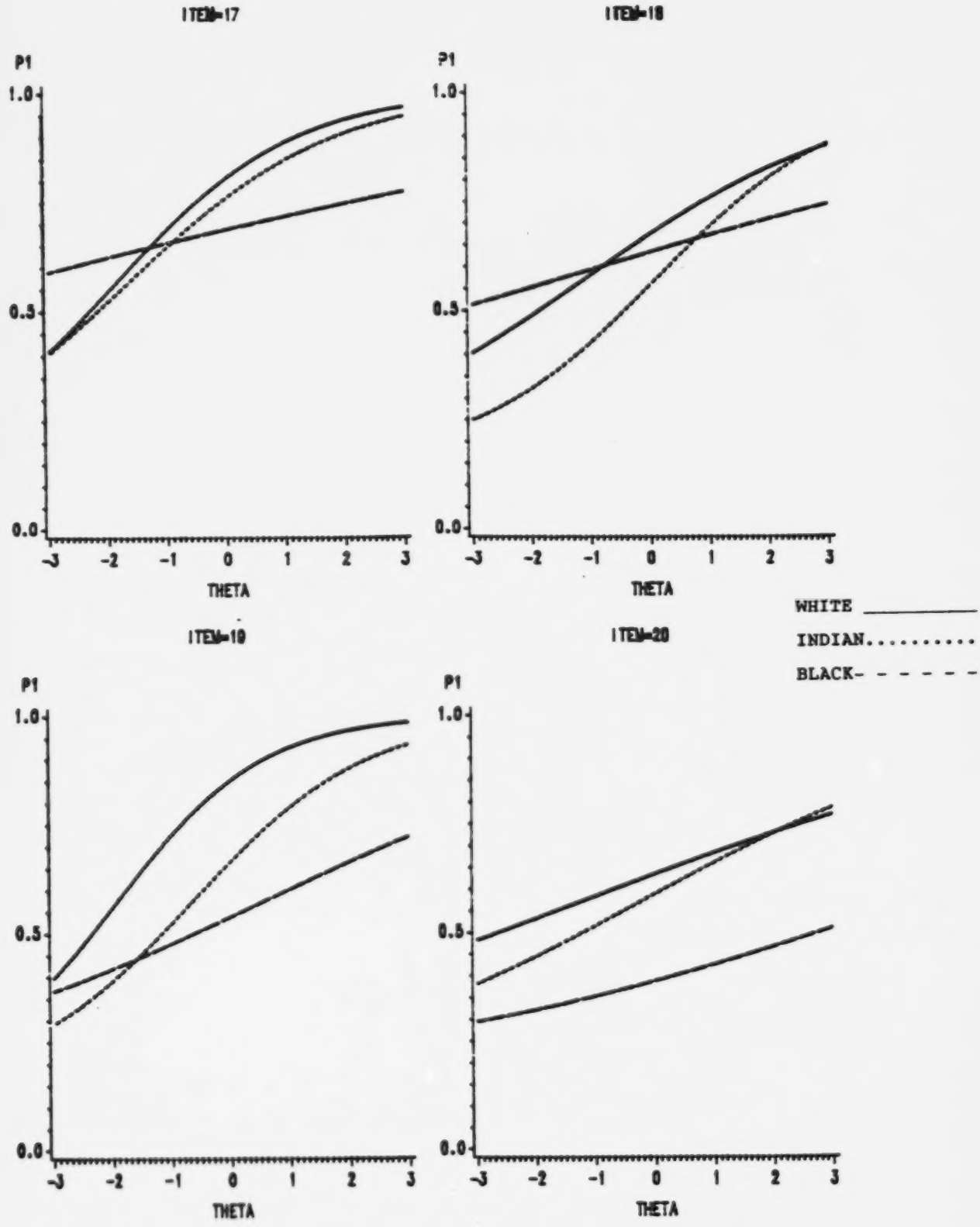


FIGURE C1: JAT 1: (CONTINUED)

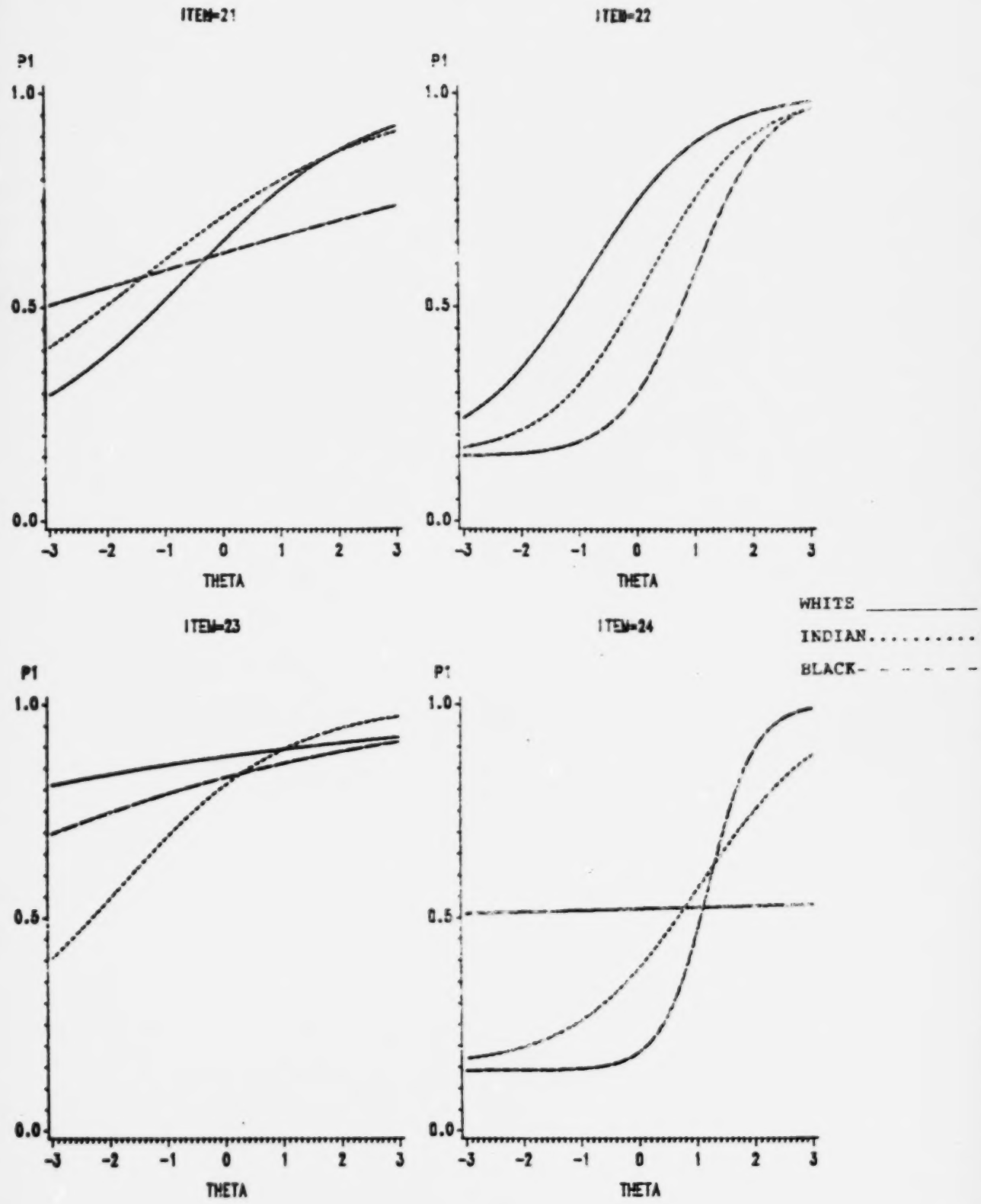


FIGURE C1: JAT 1: (CONTINUED)

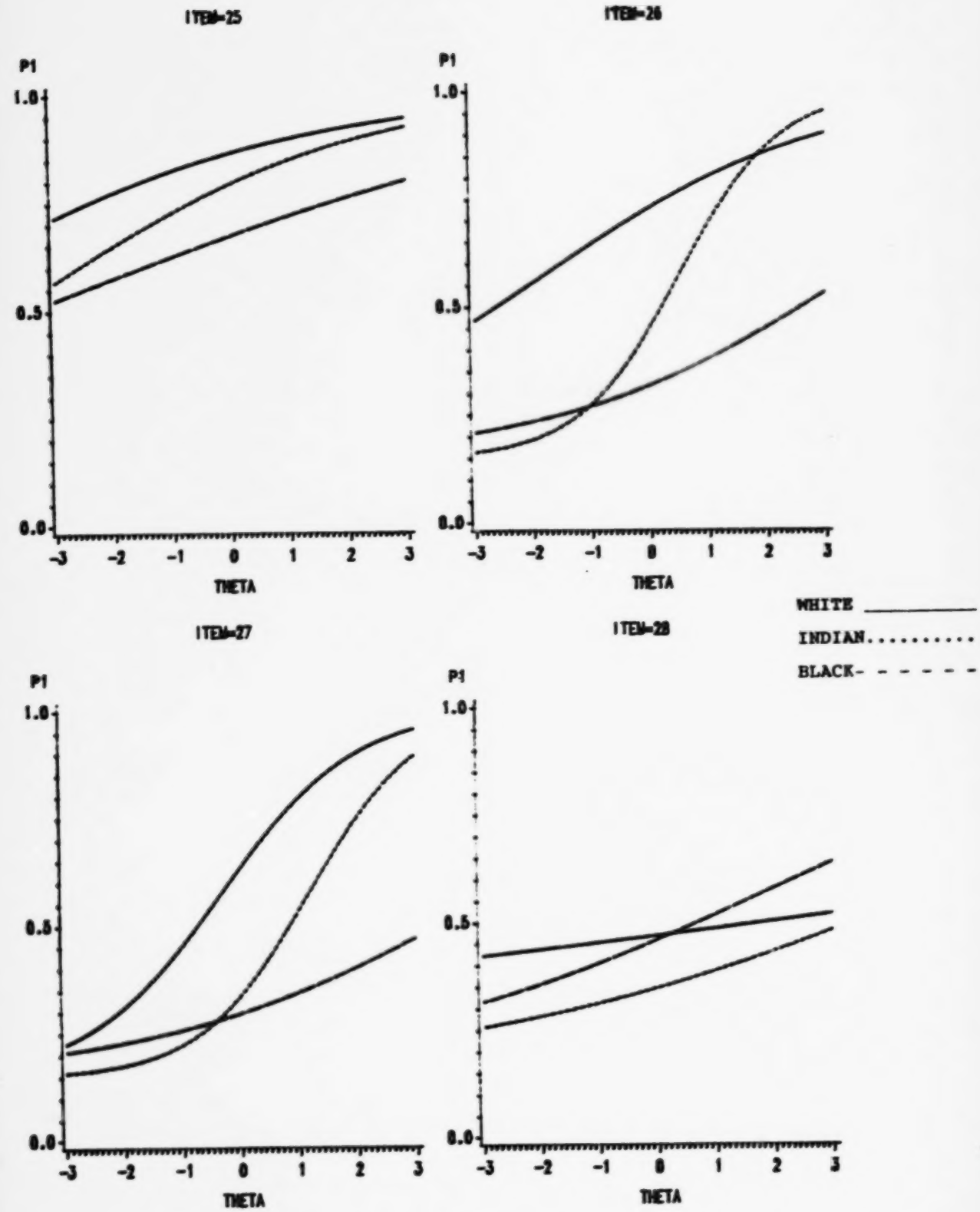


FIGURE C1: JAT 1: (CONTINUED)

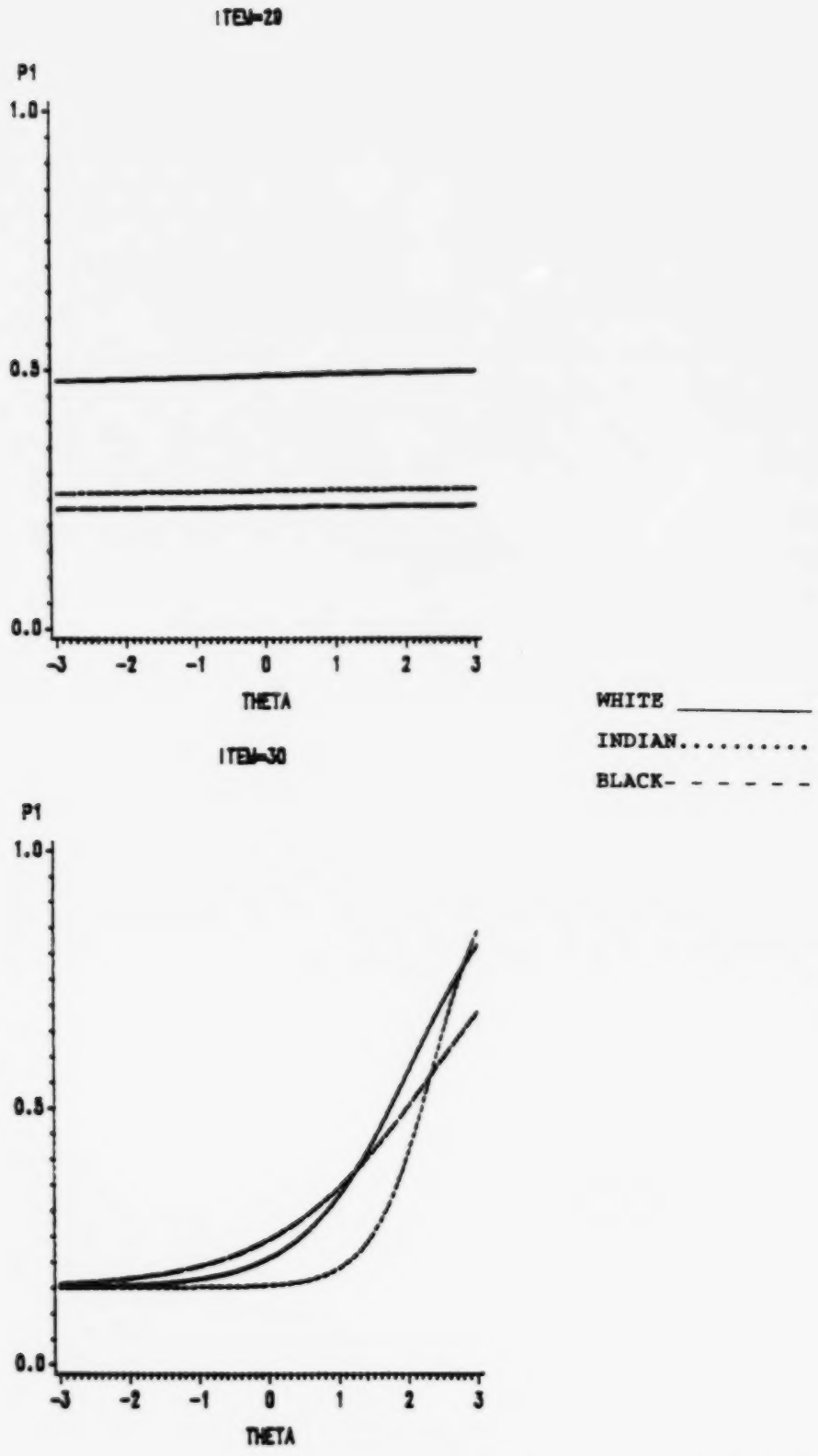


FIGURE C1: JAT 1: (CONTINUED)

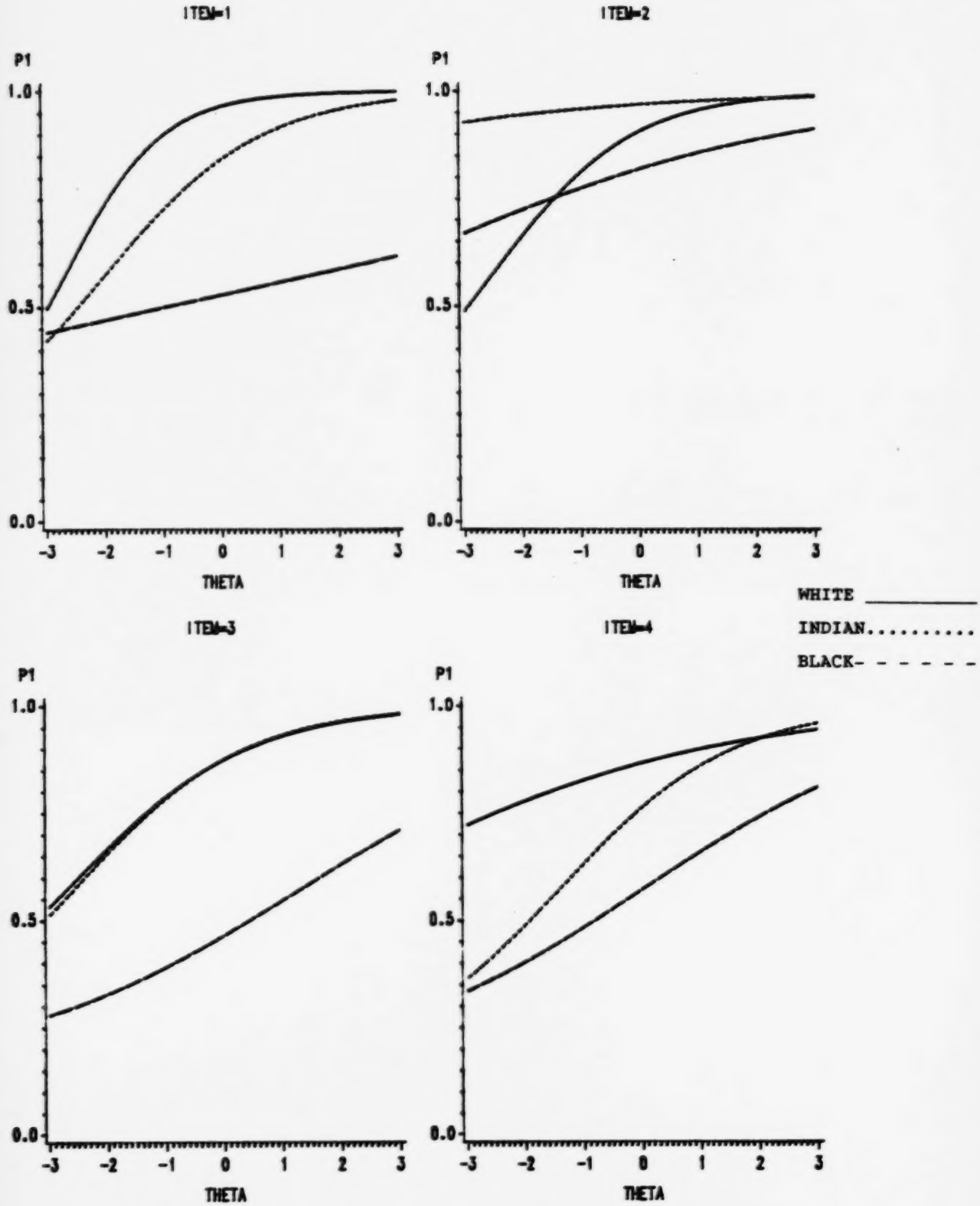


FIGURE C 2: JAT 2: ITEM CHARACTERISTIC CURVES OF THE ITEMS FOR WHITE, INDIAN AND BLACK PUPILS

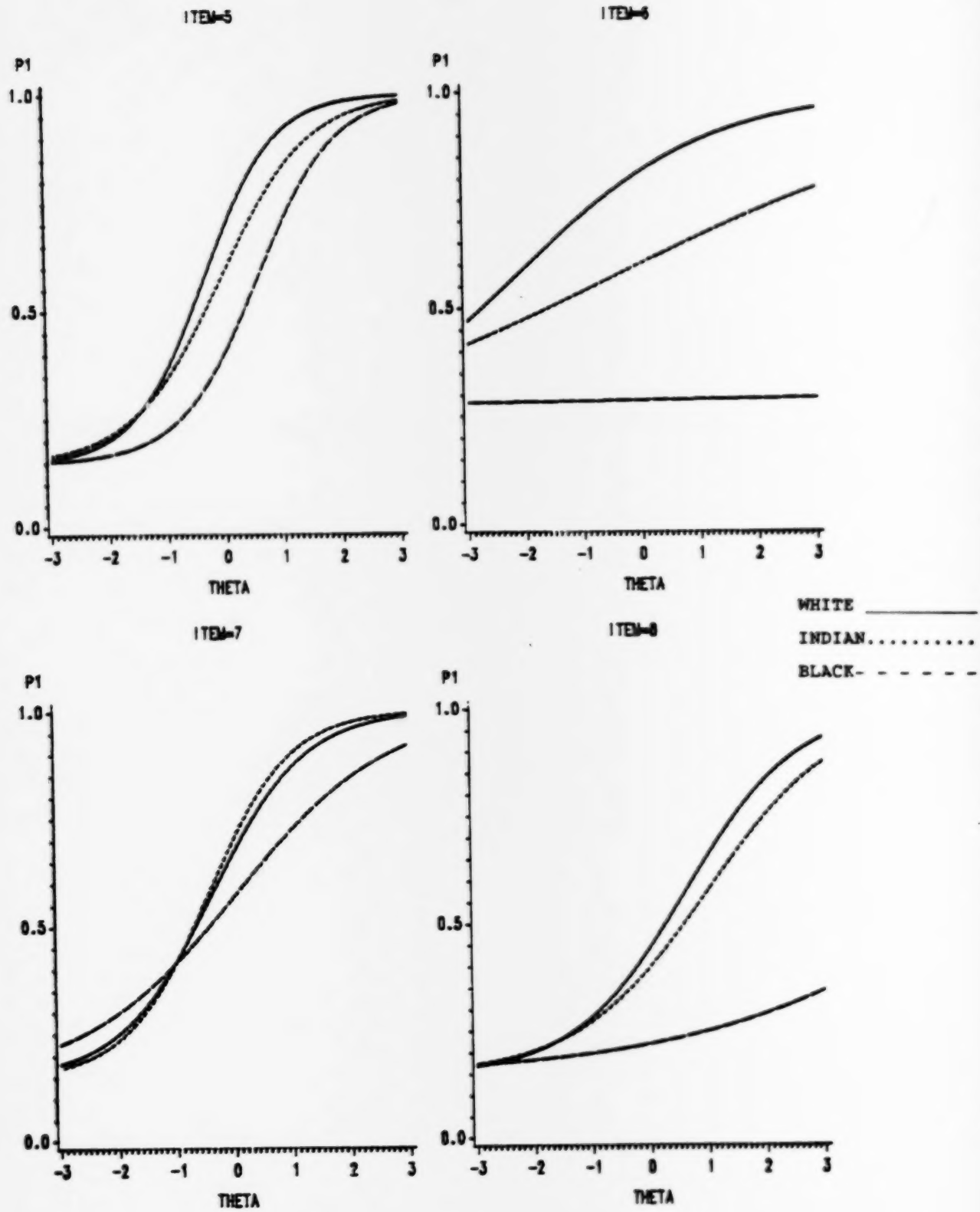


FIGURE C2: JAT 2: (CONTINUED)

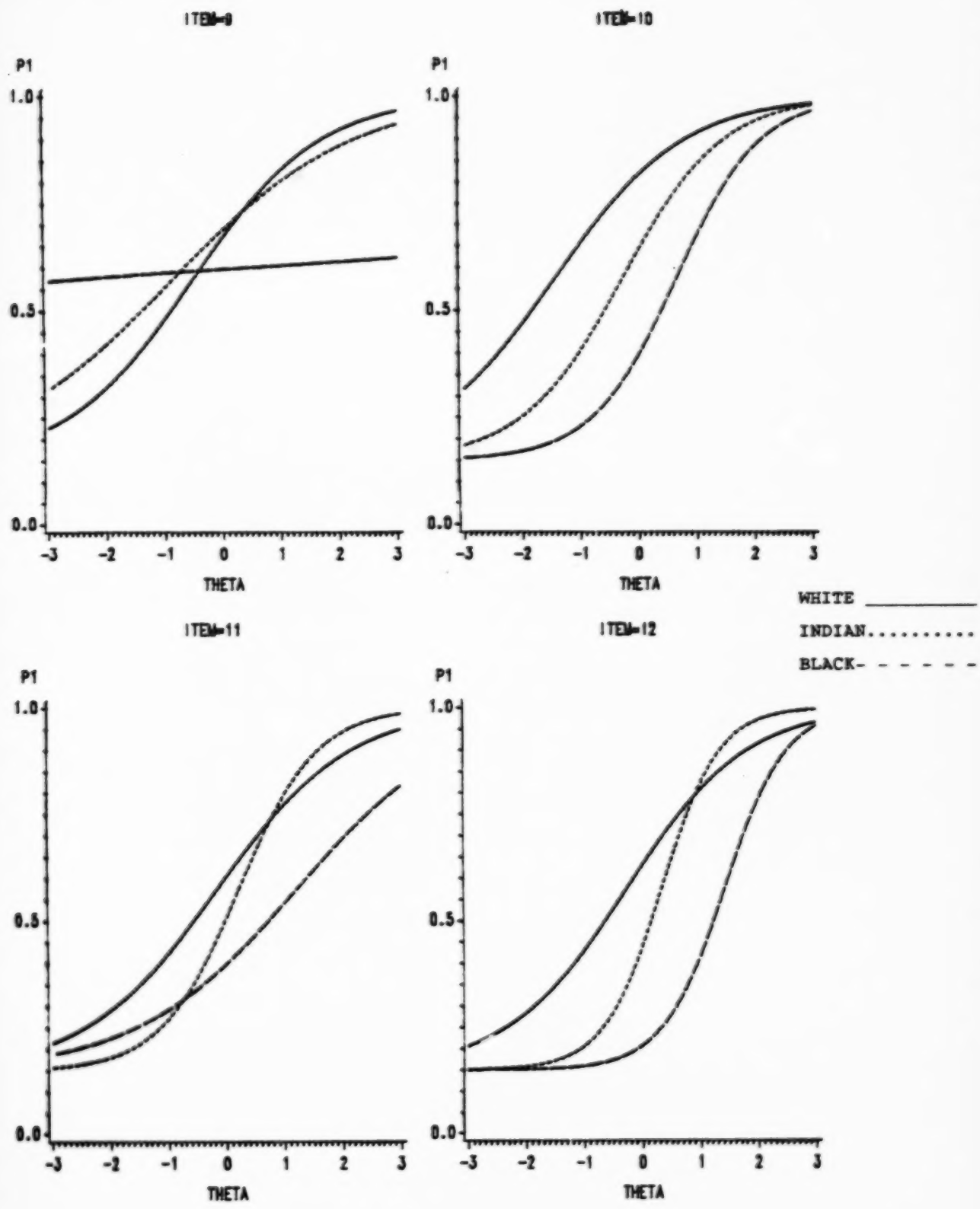


FIGURE C2: JAT 2: (CONTINUED)

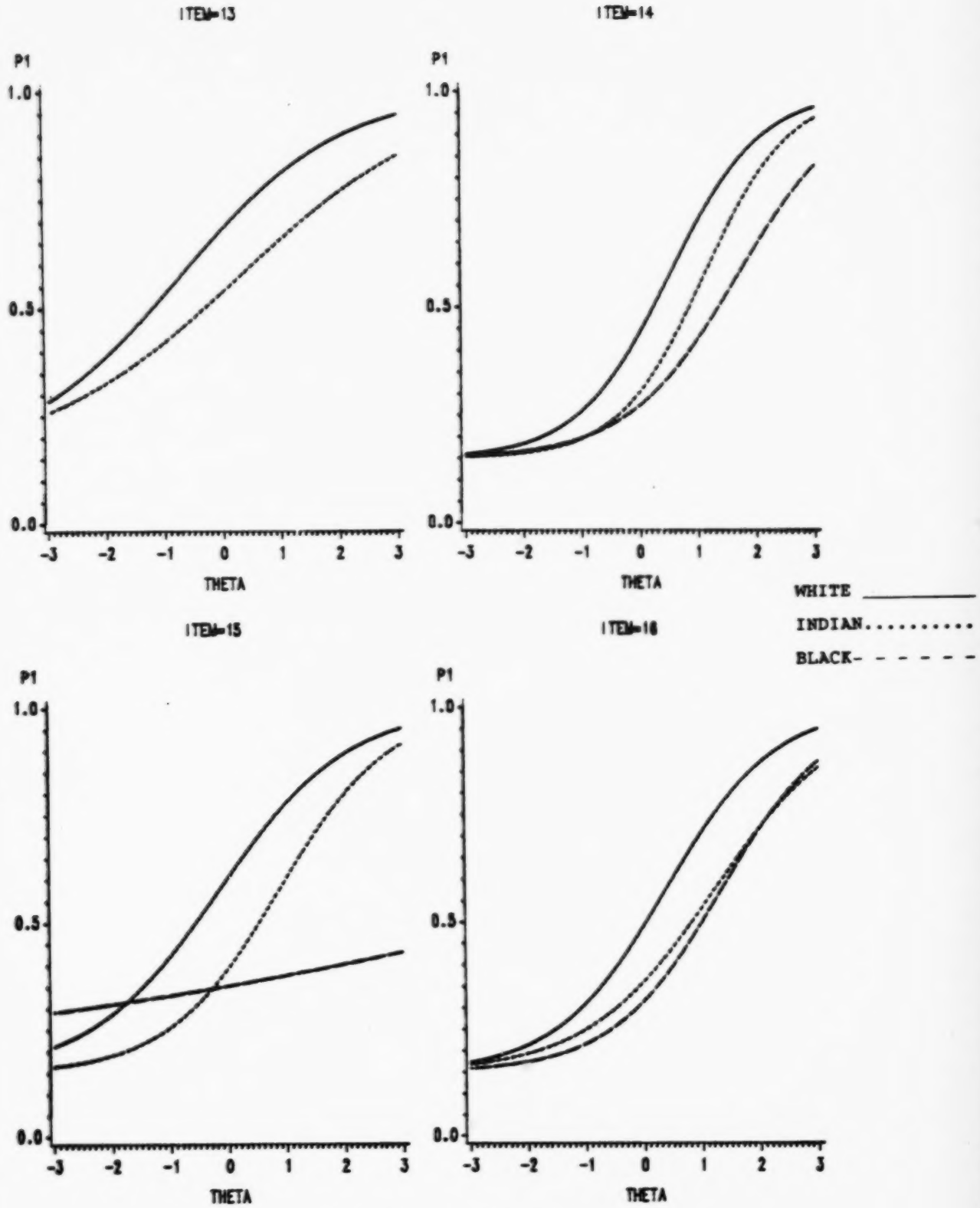


FIGURE C2: JAT 2: (CONTINUED)

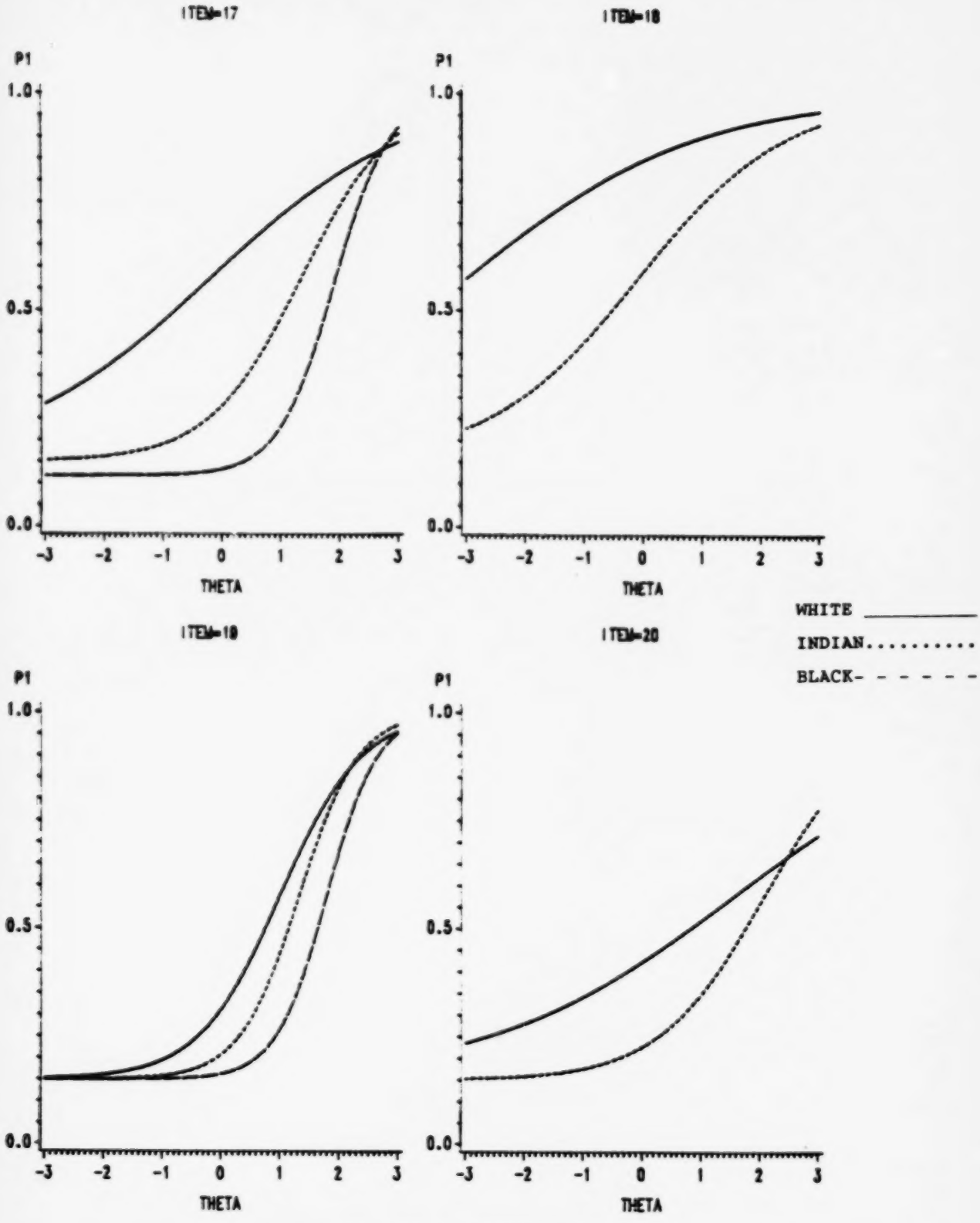


FIGURE C2: JAT 2: (CONTINUED)

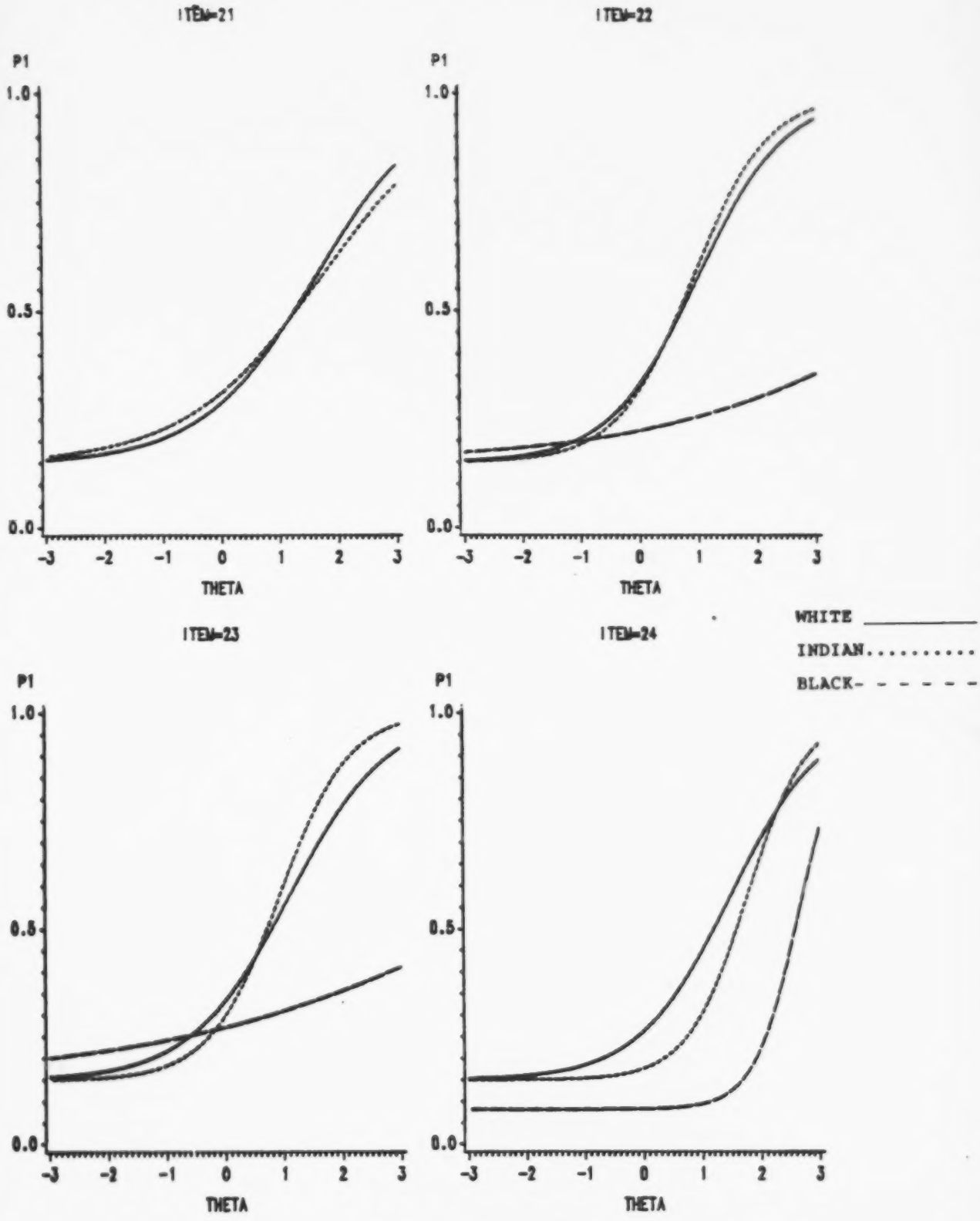


FIGURE C2: JAT 2: (CONTINUED)

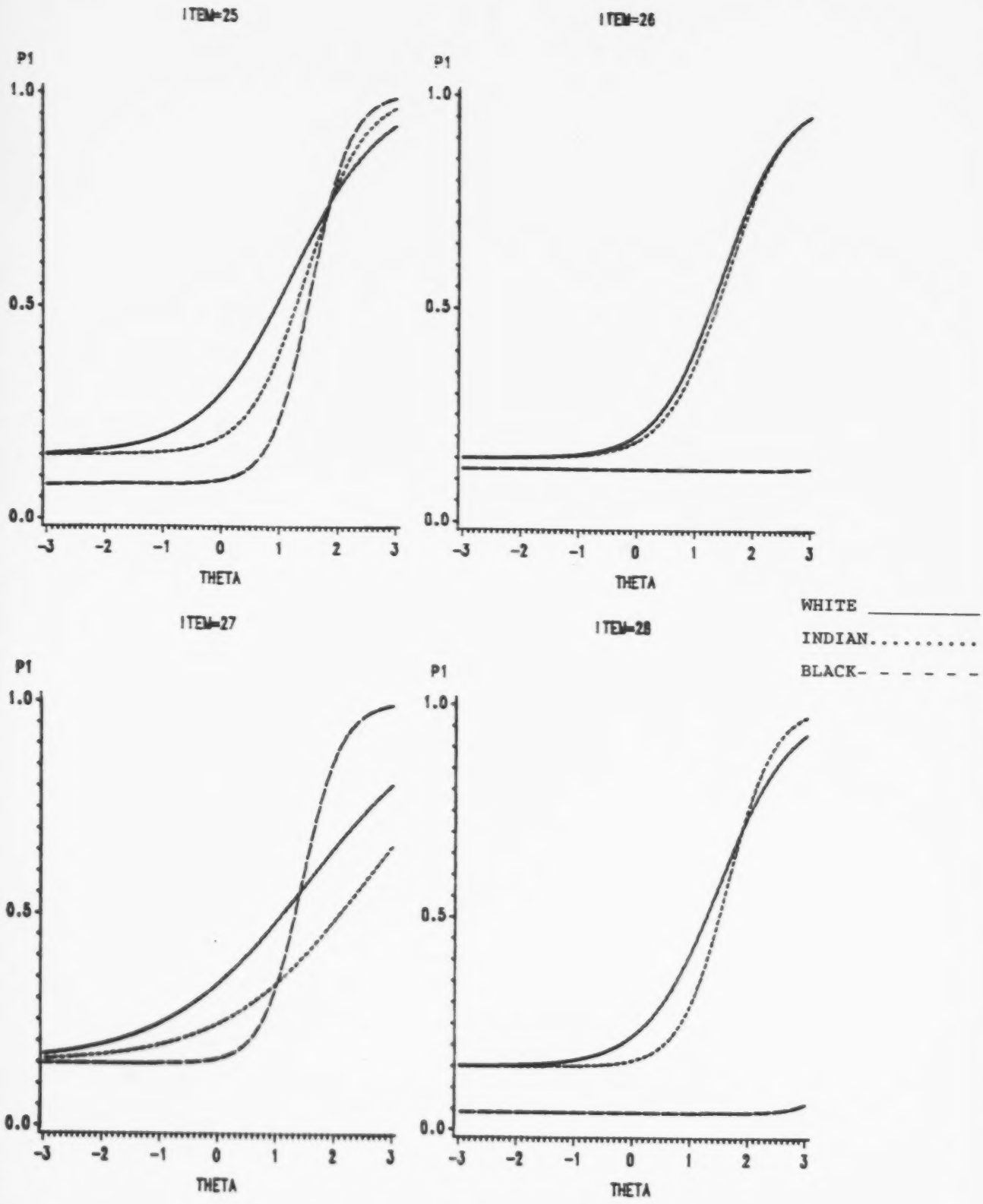


FIGURE C2: JAT 2: (CONTINUED)

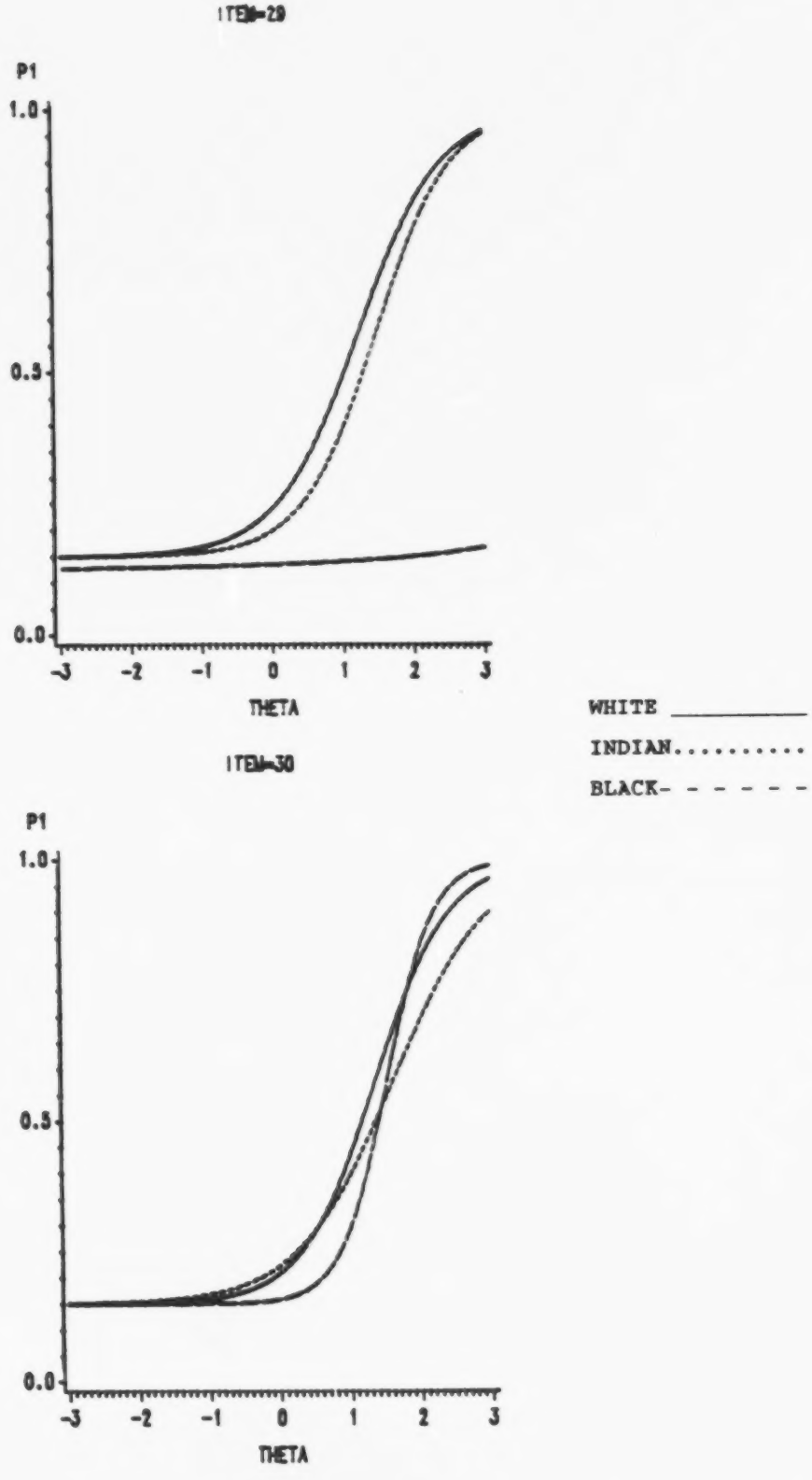


FIGURE C2: JAT 2: (CONTINUED)

HUMAN SCIENCES RESEARCH COUNCIL
RAAD VIR GEESTESWETENSKAPLIKE NAVORSING

Dr J G Garbers
President

Dr H C Marais
Deputy President
Adjunk-president

Dr K F Mauer
Vice-President
Vise-president

Mr/Mnr E E Garner
Man.: Finance and Corporate Services
Best.: Finansies en Korporatiewe Dienste

The HSRC endeavours to contribute, through all its functions, towards improving the quality of life of the inhabitants of South Africa

Functions of the HSRC

- determining research priorities
- undertaking, promoting, supporting and co-ordinating human sciences research
- disseminating research findings
- evaluating and promoting the implementation of research findings

Research Institutes

- Communication (ICOMM)
- Education (IER)
- History (IHR)
- Labour Economics (ILER)
- Language and the Arts (IRLA)
- Personnel (NIPR)
- Psychology and Edumetrics (IPER)
- Research Development (IRD)
- Sociology and Demography (ISODEM)
- Statistics (ISR)

Other

- Bureau for Liaison and Marketing (BLM)
- Bureau for Support Services (BSS)
- Group for Methodology and Technology (GMT)

Head office

Private Bag X41, Pretoria 0001
Republic of South Africa
Telegrams RAGEN
Telephone (012) 202 9111
Telex 3 20893 SA
Fax (012) 265362

Other addresses

National Institute for Personnel Research (NIPR)

P.O. Box 32410, Braamfontein 2017
Republic of South Africa
Telegrams NAVORSPERS
Telephone (011) 339 4451
Telex 4 25459 SA
Fax (011) 403 2353

Western Cape Regional Office

Private Bag X5, Roggebaai 8012
Tel. (021) 419 2572/3/4/5
Telex 5 22260 SA
Fax (021) 4196766

Natal Regional Office

P.O. Box 17302, Congella 4013
Tel. (031) 815 970
Telex 6 28567 SA
Fax (031) 812509

Eastern Cape Regional Office

P.O. Box 1124, Port Elizabeth 6000
Tel. (041) 532113
Telex 2 45183 SA
Fax (041) 532325

Die RGN poog om deur al sy funksies 'n bydrae te maak om die lewensgehalte van die inwoners van Suid-Afrika te verhoog

Funksies van die RGN

- bepaling van navorsingsprioriteite
- uitvoering, bevordering, ondersteuning en koördinering van geesteswetenskaplike navorsing
- disseminasie van navorsingsbevindinge
- evaluering en bevordering van die implementering van navorsingsbevindinge

Navorsingsinstitute

- Kommunikasie (IKOMM)
- Opvoedkunde (ION)
- Geskiedenis (IGN)
- Arbeidsekonomie (INAN)
- Taal en Kunste (INTAK)
- Personeel (NIPN)
- Psigologie en Edumetrika (IPEN)
- Navorsingsontwikkeling (INO)
- Sosiologie en Demografie (ISODEM)
- Statistiek (ISN)

Ander

- Buro vir Skakeling en Bemarking (BSB)
- Buro vir Ondersteunende Dienste (BOD)
- Groep vir Metodologie en Tegnologie (GMT)

Hoofkantoor

Privaatsak X41, Pretoria 0001
Republiek van Suid-Afrika
Telegramme RAGEN
Telefoon (012) 202 9111
Teleks 3 20893 SA
Faks (012) 265362

Ander adresse

Nasionale Instituut vir Personeelnavorsing (NIPN)

Posbus 32410, Braamfontein 2017
Republiek van Suid-Afrika
Telegramme NAVORSPERS
Telefoon (011) 339 4451
Teleks 4 25459 SA
Faks (011) 403 2353

Wes-Kaaplandse Streekkantoor

Privaatsak X5, Roggebaai 8012
Tel. (021) 419 2572/3/4/5
Teleks 5 22260 SA
Faks (021) 4196766

Natalse Streekkantoor

Posbus 17302, Congella 4013
Tel. (031) 815 970
Teleks 6 28567 SA
Faks (031) 812509

Oos-Kaaplandse Streekkantoor

Posbus 1124, Port Elizabeth 6000
Tel. (041) 532113
Teleks 2 45183 SA
Faks (041) 532325

ISBN 0 7969 0721 8

R16.00

1.0